GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2006, 17:01:08 ; Search time 1727 Seconds
(without alignments)
1974.874 Million cell updates/sec

Title: US-10-677-777-5
Perfect score: 60
Sequence: 1 ctcgggaaagggaaaccgaa.....ggaaaccgaaactgaagccc 60

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	60	100.0	60	6	CQ815109	CQ815109 Sequence
2	60	100.0	4412	6	CQ815107	CQ815107 Sequence
3	60	100.0	5849	6	CQ815105	CQ815105 Sequence
c 4	33	55.0	186798	14	AC114758	AC114758 Homo sapi
c 5	32.2	53.7	140544	14	AC011798	AC011798 Homo sapi
6	32.2	53.7	152181	14	AC027253	AC027253 Homo sapi
c 7	32.2	53.7	193047	8	AC147069	AC147069 Pan trogl
c 8	31.6	52.7	36858	3	CR626857	CR626857 Rice Clus
c 9	31.6	52.7	200418	8	AC074138	AC074138 Homo sapi
c 10	31	51.7	466	6	AR503827	AR503827 Sequence
c 11	31	51.7	466	6	AR519109	AR519109 Sequence
c 12	31	51.7	26002	14	AC014250	AC014250 Drosophil
13	31	51.7	175053	2	AC010581	AC010581 Drosophil
14	31	51.7	243348	2	AE003683	AE003683 Drosophil
c 15	30.6	51.0	180889	14	AC079986	AC079986 Homo sapi
16	30.6	51.0	188530	14	AC079547	AC079547 Mus muscu
c 17	30.6	51.0	192724	14	AC073825	AC073825 Mus muscu
18	30.6	51.0	193643	8	AC018644	AC018644 Homo sapi
19	30.4	50.7	143589	9	AC124745	AC124745 Mus muscu
c 20	30.4	50.7	165349	9	AC134870	AC134870 Mus muscu
21	30.4	50.7	208952	9	AC151994	AC151994 Mus muscu
22	30	50.0	33	6	AR065099	AR065099 Sequence
23	30	50.0	33	6	I35811	I35811 Sequence 8
24	30	50.0	33	6	AR269270	AR269270 Sequence
25	30	50.0	34	6	AR411126	AR411126 Sequence
26	30	50.0	40	6	I15313	I15313 Sequence 14
27	30	50.0	40	6	I31725	I31725 Sequence 14
28	30	50.0	40	6	I93569	I93569 Sequence 14
29	30	50.0	223	8	HUMISGA1	M31020 Human inter
c 30	30	50.0	6061	6	AX252010	AX252010 Sequence
c 31	30	50.0	6061	6	AX346466	AX346466 Sequence
c 32	30	50.0	6061	6	AX349025	AX349025 Sequence
c 33	30	50.0	6510	6	AX251401	AX251401 Sequence
c 34	30	50.0	6510	6	AX251810	AX251810 Sequence
c 35	30	50.0	6510	6	AX345296	AX345296 Sequence
c 36	30	50.0	71571	9	BX530028	BX530028 Mouse DNA
37	30	50.0	73156	9	BX842662	BX842662 Mouse DNA
c 38	30	50.0	74288	14	CR936373	CR936373 Homo sapi
c 39	30	50.0	74340	9	AC026232	AC026232 Mus muscu
40	30	50.0	89807	8	AL138875	AL138875 Human DNA
c 41	30	50.0	110000	14	AC151844_3	Continuation (4 of
c 42	30	50.0	111775	9	AC018559	AC018559 Mus muscu
43	30	50.0	115704	14	AC151865	AC151865 Ornithorh
44	30	50.0	128026	9	AC004155	AC004155 Mus muscu
45	30	50.0	128301	9	AC154721	AC154721 Mus muscu

ALIGNMENTS

RESULT 1
CQ815109
LOCUS CQ815109 60 bp DNA linear PAT 24-MAY-2004
DEFINITION Sequence 5 from Patent WO2004039990.
ACCESSION CQ815109
VERSION CQ815109.1 GI:47604199
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Tovey,M.G. and Lallemand,C.
TITLE Gene reporter assay, kit, and cells for determining the presence and/or the level of a molecule that activates signal transduction activity of a cell surface protein
JOURNAL Patent: WO 2004039990-A 5 13-MAY-2004;
Neutekbio Limited (IE)
FEATURES Location/Qualifiers
source 1. .60
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Synthetic"
ORIGIN

Query Match 100.0%; Score 60; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 5.8e-10;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60

RESULT 2
CQ815107
LOCUS CQ815107 4412 bp DNA linear PAT 24-MAY-2004
DEFINITION Sequence 3 from Patent WO2004039990.
ACCESSION CQ815107
VERSION CQ815107.1 GI:47604197
KEYWORDS .

```
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Tovey,M.G. and Lallemand,C.
  TITLE     Gene reporter assay, kit, and cells for determining the presence
            and/or the level of a molecule that activates signal transduction
            activity of a cell surface protein
  JOURNAL   Patent: WO 2004039990-A 3 13-MAY-2004;
            Neutekbio Limited (IE)
FEATURES             Location/Qualifiers
     source          1. .4412
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
     CDS             358. .1077
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAG30029.1"
                     /db_xref="GI:47604198"
                     /translation="MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT
                     LKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFK
                     DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKN
                     GIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDH
                     MVLLEFVTAAGITLGMDELYK"
ORIGIN

  Query Match             100.0%;  Score 60;  DB 6;  Length 4412;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-10;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         30 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 89


RESULT 3
CQ815105
LOCUS       CQ815105                5849 bp    DNA     linear   PAT 24-MAY-2004
DEFINITION  Sequence 1 from Patent WO2004039990.
ACCESSION   CQ815105
VERSION     CQ815105.1  GI:47604195
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Tovey,M.G. and Lallemand,C.
  TITLE     Gene reporter assay, kit, and cells for determining the presence
            and/or the level of a molecule that activates signal transduction
            activity of a cell surface protein
  JOURNAL   Patent: WO 2004039990-A 1 13-MAY-2004;
            Neutekbio Limited (IE)
FEATURES             Location/Qualifiers
     source          1. .5849
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic"
     CDS             328. .1980
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAG30028.1"
                     /db_xref="GI:47604196"
                     /translation="MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTD
                     AHIEVNITYAEYFEMSVRLAEAMKRYGLNTNHRIVVCSENSLQFFMPVLGALFIGVAV
                     APANDIYNERELLNSMNISQPTVVFVSKKGLQKILNVQKKLPIIQKIIIMDSKTDYQG
                     FQSMYTFVTSHLPPGFNEYDFVPESFDRDKTIALIMNSSGSTGLPKGVALPHRTACVR
                     FSHARDPIFGNQIIPDTAILSVVPFHHGFGMFTTLGYLICGFRVVLMYRFEEELFLRS
                     LQDYKIQSALLVPTLFSFFAKSTLIDKYDLSNLHEIASGGAPLSKEVGEAVAKRFHLP
                     GIRQGYGLTETTSAILITPEGDDKPGAVGKVVPFFEAKVVDLDTGKTLGVNQRGELCV
                     RGPMIMSGYVNNPEATNALIDKDGWLHSGDIAYWDEDEHFFIVDRLKSLIKYKGYQVA
                     PAELESILLQHPNIFDAGVAGLPDDDAGELPAAVVVLEHGKTMTEKEIVDYVASQVTT
                     AKKLRGGVVFVDEVPKGLTGKLDARKIREILIKAKKGGKSKL"
ORIGIN

  Query Match             100.0%;  Score 60;  DB 6;  Length 5849;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-10;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         38 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 97


RESULT 4
AC114758/c
LOCUS       AC114758              186798 bp    DNA     linear   HTG 11-MAR-2002
DEFINITION  Homo sapiens chromosome 4 clone RP11-339C11, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC114758
VERSION     AC114758.1  GI:19339143
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 186798)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186798)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0339C11
            ------------------------------------------------------
            ------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 176113 bases at least Q40
            Consensus quality: 178045 bases at least Q30
            Consensus quality: 180548 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 185574; sum-of-contigs
            Quality coverage: 7.26 in Q20 bases; agarose-fp
            Quality coverage: 6.82 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      224: contig of 224 bp in length
            *      225      324: gap of unknown length
            *      325    56954: contig of 56630 bp in length
            *    56955    57054: gap of unknown length
```

```
            *    57055    58456: contig of 1402 bp in length
            *    58457    58556: gap of unknown length
            *    58557    59797: contig of 1241 bp in length
            *    59798    59897: gap of unknown length
            *    59898    61436: contig of 1539 bp in length
            *    61437    61536: gap of unknown length
            *    61537    62662: contig of 1126 bp in length
            *    62663    62762: gap of unknown length
            *    62763    64174: contig of 1412 bp in length
            *    64175    64274: gap of unknown length
            *    64275    66267: contig of 1993 bp in length
            *    66268    66367: gap of unknown length
            *    66368    67478: contig of 1111 bp in length
            *    67479    67578: gap of unknown length
            *    67579   114758: contig of 47180 bp in length
            *   114759   114858: gap of unknown length
            *   114859   186798: contig of 71940 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186798
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-339C11"
     misc_feature    1. .224
                     /note="assembly_name:Contig10"
     gap             225. .324
                     /estimated_length=unknown
     misc_feature    325. .56954
                     /note="assembly_name:Contig63
                     clone_end:T7
                     vector_side:left"
     gap             56955. .57054
                     /estimated_length=unknown
     misc_feature    57055. .58456
                     /note="assembly_name:Contig44"
     gap             58457. .58556
                     /estimated_length=unknown
     misc_feature    58557. .59797
                     /note="assembly_name:Contig52"
     gap             59798. .59897
                     /estimated_length=unknown
     misc_feature    59898. .61436
                     /note="assembly_name:Contig53"
     gap             61437. .61536
                     /estimated_length=unknown
     misc_feature    61537. .62662
                     /note="assembly_name:Contig56"
     gap             62663. .62762
                     /estimated_length=unknown
     misc_feature    62763. .64174
                     /note="assembly_name:Contig57"
     gap             64175. .64274
                     /estimated_length=unknown
     misc_feature    64275. .66267
                     /note="assembly_name:Contig60"
     gap             66268. .66367
                     /estimated_length=unknown
     misc_feature    66368. .67478
                     /note="assembly_name:Contig61"
     gap             67479. .67578
                     /estimated_length=unknown
     misc_feature    67579. .114758
                     /note="assembly_name:Contig62"
     gap             114759. .114858
                     /estimated_length=unknown
     misc_feature    114859. .186798
                     /note="assembly_name:Contig64"
ORIGIN

  Query Match             55.0%;  Score 33;  DB 14;  Length 186798;
  Best Local Similarity   73.7%;  Pred. No. 1.6;
  Matches   42;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              ||||||||||||||| |||| ||| |||||||||||| |||| ||||
Db      18949 CTCGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAG 18893


RESULT 5
AC011798/c
LOCUS       AC011798              140544 bp    DNA     linear   HTG 20-FEB-2000
DEFINITION  Homo sapiens clone RP11-17M21, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC011798
VERSION     AC011798.3  GI:7008800
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 140544)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-17M21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140544)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 20, 2000 this sequence version replaced gi:6446858.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3664
                Center clone name: 17_M_21
            --------------.
            * NOTE: This record contains 172 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      814: contig of 814 bp in length
            *                    gap of unknown length
            *      815     1624: contig of 810 bp in length
            *                    gap of unknown length
            *     1625     2397: contig of 773 bp in length
            *                    gap of unknown length
            *     2398     3186: contig of 789 bp in length
            *                    gap of unknown length
            *     3187     3980: contig of 794 bp in length
```

```
*                    gap of unknown length
*       3981    4793: contig of 813 bp in length
*                    gap of unknown length
*       4794    5603: contig of 810 bp in length
*                    gap of unknown length
*       5604    6410: contig of 807 bp in length
*                    gap of unknown length
*       6411    7213: contig of 803 bp in length
*                    gap of unknown length
*       7214    7969: contig of 756 bp in length
*                    gap of unknown length
*       7970    8763: contig of 794 bp in length
*                    gap of unknown length
*       8764    9566: contig of 803 bp in length
*                    gap of unknown length
*       9567   10353: contig of 787 bp in length
*                    gap of unknown length
*      10354   11127: contig of 774 bp in length
*                    gap of unknown length
*      11128   11921: contig of 794 bp in length
*                    gap of unknown length
*      11922   12729: contig of 808 bp in length
*                    gap of unknown length
*      12730   13533: contig of 804 bp in length
*                    gap of unknown length
*      13534   14326: contig of 793 bp in length
*                    gap of unknown length
*      14327   15141: contig of 815 bp in length
*                    gap of unknown length
*      15142   15968: contig of 827 bp in length
*                    gap of unknown length
*      15969   16791: contig of 823 bp in length
*                    gap of unknown length
*      16792   17496: contig of 705 bp in length
*                    gap of unknown length
*      17497   18269: contig of 773 bp in length
*                    gap of unknown length
*      18270   19033: contig of 764 bp in length
*                    gap of unknown length
*      19034   19817: contig of 784 bp in length
*                    gap of unknown length
*      19818   20572: contig of 755 bp in length
*                    gap of unknown length
*      20573   21387: contig of 815 bp in length
*                    gap of unknown length
*      21388   22199: contig of 812 bp in length
*                    gap of unknown length
*      22200   23009: contig of 810 bp in length
*                    gap of unknown length
*      23010   23816: contig of 807 bp in length
*                    gap of unknown length
*      23817   24598: contig of 782 bp in length
*                    gap of unknown length
*      24599   25425: contig of 827 bp in length
*                    gap of unknown length
*      25426   26222: contig of 797 bp in length
*                    gap of unknown length
*      26223   26992: contig of 770 bp in length
*                    gap of unknown length
*      26993   27802: contig of 810 bp in length
*                    gap of unknown length
*      27803   28598: contig of 796 bp in length
*                    gap of unknown length
*      28599   29366: contig of 768 bp in length
*                    gap of unknown length
*      29367   30189: contig of 823 bp in length
*                    gap of unknown length
*      30190   30937: contig of 748 bp in length
*                    gap of unknown length
*      30938   31755: contig of 818 bp in length
*                    gap of unknown length
*      31756   32570: contig of 815 bp in length
*                    gap of unknown length
*      32571   33391: contig of 821 bp in length
*                    gap of unknown length
*      33392   34200: contig of 809 bp in length
*                    gap of unknown length
*      34201   34998: contig of 798 bp in length
*                    gap of unknown length
*      34999   35770: contig of 772 bp in length
*                    gap of unknown length
*      35771   36547: contig of 777 bp in length
*                    gap of unknown length
*      36548   37335: contig of 788 bp in length
*                    gap of unknown length
*      37336   38118: contig of 783 bp in length
*                    gap of unknown length
*      38119   38882: contig of 764 bp in length
*                    gap of unknown length
*      38883   39696: contig of 814 bp in length
*                    gap of unknown length
*      39697   40507: contig of 811 bp in length
*                    gap of unknown length
*      40508   41316: contig of 809 bp in length
*                    gap of unknown length
*      41317   42138: contig of 822 bp in length
*                    gap of unknown length
*      42139   42902: contig of 764 bp in length
*                    gap of unknown length
*      42903   43691: contig of 789 bp in length
*                    gap of unknown length
*      43692   44471: contig of 780 bp in length
*                    gap of unknown length
*      44472   45256: contig of 785 bp in length
*                    gap of unknown length
*      45257   46049: contig of 793 bp in length
*                    gap of unknown length
*      46050   46865: contig of 816 bp in length
*                    gap of unknown length
*      46866   47684: contig of 819 bp in length
*                    gap of unknown length
*      47685   48474: contig of 790 bp in length
*                    gap of unknown length
*      48475   49300: contig of 826 bp in length
*                    gap of unknown length
*      49301   50115: contig of 815 bp in length
*                    gap of unknown length
*      50116   50875: contig of 760 bp in length
*                    gap of unknown length
*      50876   51690: contig of 815 bp in length
*                    gap of unknown length
*      51691   52491: contig of 801 bp in length
*                    gap of unknown length
*      52492   53257: contig of 766 bp in length
*                    gap of unknown length
*      53258   54009: contig of 752 bp in length
*                    gap of unknown length
*      54010   54773: contig of 764 bp in length
*                    gap of unknown length
*      54774   55564: contig of 791 bp in length
*                    gap of unknown length
*      55565   56349: contig of 785 bp in length
*                    gap of unknown length
*      56350   57162: contig of 813 bp in length
*                    gap of unknown length

  Query Match           53.7%;  Score 32.2;  DB 14;  Length 140544;
  Best Local Similarity  75.5%;  Pred. No. 2.9;
  Matches   40;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              |||||||||||||  |||||  |||       |  ||||||||||   ||||  ||||
Db       5064 GGGAAAGGGAAACGGAAACGGAAAGTGACAGAGAAAGGGAAAGGGAAAGTGAA 5012

RESULT 6
```

```
AC027253
LOCUS       AC027253              152181 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 7 clone RP11-37J11 map 7, WORKING DRAFT
            SEQUENCE, 25 unordered pieces.
ACCESSION   AC027253
VERSION     AC027253.2  GI:8516121
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 152181)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 7, clone RP11-37J11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152181)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 152181)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 14, 2000 this sequence version replaced gi:7331623.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8929
                Center clone name: 37_J_11
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 134542 bases at least Q40
                Consensus quality: 143434 bases at least Q30
                Consensus quality: 147452 bases at least Q20
                Insert size: 144000; agarose-fp
                Insert size: 149781; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1307: contig of 1307 bp in length
            *     1308     1407: gap of 100 bp
            *     1408     2930: contig of 1523 bp in length
            *     2931     3030: gap of 100 bp
            *     3031     5024: contig of 1994 bp in length
            *     5025     5124: gap of 100 bp
            *     5125     7827: contig of 2703 bp in length
            *     7828     7927: gap of 100 bp
            *     7928    11667: contig of 3740 bp in length
            *    11668    11767: gap of 100 bp
            *    11768    15494: contig of 3727 bp in length
            *    15495    15594: gap of 100 bp
            *    15595    18512: contig of 2918 bp in length
            *    18513    18612: gap of 100 bp
            *    18613    21649: contig of 3037 bp in length
            *    21650    21749: gap of 100 bp
            *    21750    25120: contig of 3371 bp in length
            *    25121    25220: gap of 100 bp
            *    25221    29007: contig of 3787 bp in length
            *    29008    29107: gap of 100 bp
            *    29108    33722: contig of 4615 bp in length
            *    33723    33822: gap of 100 bp
            *    33823    36660: contig of 2838 bp in length
            *    36661    36760: gap of 100 bp
            *    36761    41284: contig of 4524 bp in length
            *    41285    41384: gap of 100 bp
            *    41385    47138: contig of 5754 bp in length
            *    47139    47238: gap of 100 bp
            *    47239    53124: contig of 5886 bp in length
            *    53125    53224: gap of 100 bp
            *    53225    59057: contig of 5833 bp in length
            *    59058    59157: gap of 100 bp
            *    59158    64096: contig of 4939 bp in length
            *    64097    64196: gap of 100 bp
            *    64197    71094: contig of 6898 bp in length
            *    71095    71194: gap of 100 bp
            *    71195    77922: contig of 6728 bp in length
            *    77923    78022: gap of 100 bp
            *    78023    90281: contig of 12259 bp in length
            *    90282    90381: gap of 100 bp
            *    90382   100336: contig of 9955 bp in length
            *   100337   100436: gap of 100 bp
            *   100437   111540: contig of 11104 bp in length
            *   111541   111640: gap of 100 bp
            *   111641   122074: contig of 10434 bp in length
            *   122075   122174: gap of 100 bp
            *   122175   134879: contig of 12705 bp in length
            *   134880   134979: gap of 100 bp
```

```
          *   134980   152181: contig of 17202 bp in length.
FEATURES             Location/Qualifiers
     source          1. .152181
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-37J11"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1307
                     /note="assembly_fragment"
     gap             1308. .1407
                     /estimated_length=100
     misc_feature    1408. .2930
                     /note="assembly_fragment"
     gap             2931. .3030
                     /estimated_length=100
     misc_feature    3031. .5024
                     /note="assembly_fragment"
     gap             5025. .5124
                     /estimated_length=100
     misc_feature    5125. .7827
                     /note="assembly_fragment"
     gap             7828. .7927
                     /estimated_length=100
     misc_feature    7928. .11667
                     /note="assembly_fragment"
     gap             11668. .11767
                     /estimated_length=100
     misc_feature    11768. .15494
                     /note="assembly_fragment"
     gap             15495. .15594
                     /estimated_length=100
     misc_feature    15595. .18512
                     /note="assembly_fragment"
     gap             18513. .18612
                     /estimated_length=100
     misc_feature    18613. .21649
                     /note="assembly_fragment"
     gap             21650. .21749
                     /estimated_length=100
     misc_feature    21750. .25120
                     /note="assembly_fragment"
     gap             25121. .25220
                     /estimated_length=100
     misc_feature    25221. .29007
                     /note="assembly_fragment"
     gap             29008. .29107
                     /estimated_length=100
     misc_feature    29108. .33722
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     gap             33723. .33822

  Query Match             53.7%;  Score 32.2;  DB 14;  Length 152181;
  Best Local Similarity   75.5%;  Pred. No. 2.9;
  Matches   40;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              ||||||||||||| ||||| |||        | ||||||||||  |||| ||||
Db     111439 GGGAAAGGGAAACGGAAACGGAAAGTGACAGAGAAAGGGAAAGGGAAAGTGAA 111491


RESULT 7
AC147069/c
LOCUS       AC147069              193047 bp    DNA     linear   PRI 19-MAY-2004
DEFINITION  Pan troglodytes BAC clone RP43-165H20 from 7, complete sequence.
ACCESSION   AC147069
VERSION     AC147069.2  GI:47131405
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 193047)
  AUTHORS   Berghoff,A., Cotton,M. and Meyer,R.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-165H20
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 193047)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 193047)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2004) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 193047)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2004) Washington University School of Medicine,
            Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO
            63108, USA
COMMENT     On May 12, 2004 this sequence version replaced gi:38154101.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_PT165H20
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
            clone and detailed information can be obtained from ResGen
            (http://www.resgen.com) or Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .193047
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /map="7"
                     /clone="RP43-165H20"
                     /clone_lib="RPCI-43"
ORIGIN

  Query Match             53.7%;  Score 32.2;  DB 8;  Length 193047;
  Best Local Similarity   75.5%;  Pred. No. 3;
```

```
  Matches   40;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              |||||||||||||| ||||| |||       | ||||||||||  |||| ||||
Db      77563 GGGAAAGGGAAACGGAAACAGAAAGTGACAGAGAAAGGGAAAGGGAAAGTGAA 77511


RESULT 8
CR626857/c
LOCUS       CR626857               36858 bp    DNA     linear   ENV 17-APR-2005
DEFINITION  Rice Cluster I (RC-I), uncultured archeon clone 3D10 complete
            sequence.
ACCESSION   CR626857
VERSION     CR626857.1  GI:56295558
KEYWORDS    rRNA operon.
SOURCE      uncultured archaeon
  ORGANISM  uncultured archaeon
            Archaea; environmental samples.
REFERENCE   1  (bases 1 to 36858)
  AUTHORS   Erkel,C., Kemnitz,D., Ricke,P., Kube,M., Chin,K.J., Dedysh,S.,
            Reinhardt,R., Conrad,R. and Liesack,W.
  TITLE     Retrieval of first genome data for rice cluster I methanogens by a
            combined approach of cultivation and molecular ecology techniques
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 36858)
  AUTHORS   Kube,M., Liesack,W., Erkel,C., Beck,A., Sontag,M. and Reinhardt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Max Planck Institute for Molecular
            Genetics, proScience Ihnestrasse 73, D-14195 Berlin, Germany
COMMENT     This project was carried out by
            *Max Planck Institute for Terrestrical Microbiology, Marburg,
            Germany; *Max Planck Institute for Molecular Genetics, Berlin,
            Germany; -------------- Genome Center
                 Center: Max Planck Institute for Molecular Genetics
                 Center code: MPIMG
            -------------- Summary Statistics
                 Sequencing vector: pUC19; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; allregions were covered by at least
            one plasmid Sequence.
            --------------.
FEATURES             Location/Qualifiers
     source          1. .36858
                     /organism="uncultured archaeon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:115547"
                     /clone="3D10mr4"
                     /note="clone of a fosmid library from an enriched
                     methanogenic consortium (Rice Cluster I) from rice field
                     soil"
     gene            complement(256. .393)
                     /gene="5S ribosomal RNA"
     rRNA            complement(256. .393)
                     /gene="5S ribosomal RNA"
                     /product="5S ribosomal RNA"
     gene            complement(546. .3469)
                     /gene="23S ribosomal RNA"
     rRNA            complement(546. .3469)
                     /gene="23S ribosomal RNA"
                     /product="23S ribosomal RNA"
     gene            complement(3814. .5288)
                     /gene="16S ribosomal RNA"
     rRNA            complement(3814. .5288)
                     /gene="16S ribosomal RNA"
                     /product="16S ribosomal RNA"
     gene            complement(5971. .7722)
                     /gene="pyK"
                     /locus_tag="orf1"
     CDS             complement(5971. .7722)
                     /gene="pyK"
                     /locus_tag="orf1"
                     /EC_number="2.7.1.40"
                     /note="IPR001697"
                     /codon_start=1
                     /transl_table=11
                     /product="pyruvate kinase (PyK)"
                     /protein_id="CAH04801.1"
                     /db_xref="GI:56295559"
                     /db_xref="GOA:Q5NW05"
                     /db_xref="InterPro:IPR001697"
                     /db_xref="InterPro:IPR008279"
                     /db_xref="UniProt/TrEMBL:Q5NW05"
                     /translation="MRKTKIVCTIGPACDSQDMLEKLAVAGMNVARLNMSHADHEHTV
                     QTINNIRMVSEAIGKPIGILMDLQGPKIRVGTLQQPANLKPGGTFTLTTRDVPGDSQE
                     VNVPFKELPQSVSTGQTLLLDDGLIELKVDAVTETDIRTKVVRGGELKSKKGINLPQS
                     TIRIPSITEKDVRDLEFGIEHEVDMIAMSFVRKPQDVLDLRKKIEDNDSDIPIISKIE
                     KHEAVKNIDGIIDVVDGVMVARGDLGIEIPMAEVPIVQKMIISKCIARGIPVITATQM
                     LDSMIRNPIPTRAEATDVANAVFDGTDALMLSGETAFGEYPVKAVETMARIAKYTEES
                     TYYKHAIAAKAPKPSLSMTDAVAQSTTESARVLKAQAIITATQTGYSARKVSKYRPQL
                     PILAVTNDPKVVNRLTLSWAVLPLLIGSPGNLDELIQDSVDACLQKGYVKNGDLVVIT
                     AGVMTGIPGGTNIMKIHVVAKELARGIGVGKDIVKGVARVIKSPSEFTNIRSGDILVI
                     KEVDVDHIDDVKPARAIISEETGLTSYSAIIGRELNIPVVVGIKNATRIIQDGSTITV
                     DSVRGMVYEGSINLPGE"
     gene            7890. .8828
                     /gene="cpsN"
                     /locus_tag="orf2"
     CDS             7890. .8828
                     /gene="cpsN"
                     /locus_tag="orf2"
                     /EC_number="4.2.1.46"
                     /note="IPR001509"
                     /codon_start=1
                     /transl_table=11
                     /product="dtdp-glucose 4,6-dehydratase"
                     /protein_id="CAH04802.1"
                     /db_xref="GI:56295560"
                     /db_xref="GOA:Q5NW04"
                     /db_xref="UniProt/TrEMBL:Q5NW04"
                     /translation="MKCIVTGGAGFIGSHLTDRLLEEGHEVTVVDNLSGGQFRFIEHH
                     VTNPEFSFVNEDLARDGLISSAFEGADMVYHLAANPDVRSGVTDTRTPLTQNTIATFN
                     VLESMRAAGVRKIAFTSTSTVYGEAEVIPTPENYGPLMPISLYGASKLACEAMISAYC
                     HTFDMQSWIYRFANIIGSRGTHGVIFDFIGRLRKDPSKLTILGNGRQSKSYLHISDCV
                     DGMLFAVRNSNAPVNIFNIGSDDRFDVTGIARAVASEMGLENVEFEYTGGDRGWKGDV
                     PFMTLSIAKLKALGWKPVHNSEESVRLCVRELLREV"
     gene            8825. .9604
                     /locus_tag="orf3"
     CDS             8825. .9604
                     /locus_tag="orf3"
                     /note="COG1718"
                     /codon_start=1
                     /transl_table=11
                     /product="serine/threonine protein kinase"
                     /protein_id="CAH04803.1"
                     /db_xref="GI:56295561"
                     /db_xref="GOA:Q5NW03"
                     /db_xref="InterPro:IPR000687"
                     /db_xref="InterPro:IPR000719"
                     /db_xref="InterPro:IPR008266"
                     /db_xref="UniProt/TrEMBL:Q5NW03"
                     /translation="MTLERPSKKEKRIDEKIDEYRIKIKDSEDLKLRGEVFDTATLKA
                     LYTFANKGIIKAMGGVVSTGKEGNVFHAIGAEDREIAIKIYRIATSDFRKMEDYMLGD
                     PRFANIKHTQKGIIFAWTQKEFRNLQRAAEAGVRVPAPIEADRNILIMEFIGKDGIPA
                     PRLRDVQLAEPEHIYRTIVSYMVALYQEAKLVHSDLSEFNILLYEDEPVIIDMGQSVL
                     LDHPMSREFLQRDVKNIVRYFKKLGVKCDEDSLLTEITTKK"
     gene            9609. .10154
                     /locus_tag="orf4"
     CDS             9609. .10154
                     /locus_tag="orf4"
                     /note="COG1094; IPR004087 (KHdomain)"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="RNA-binding protein"
                     /protein_id="CAH04804.1"
                     /db_xref="GI:56295562"
                     /db_xref="GOA:Q5NW02"
                     /db_xref="InterPro:IPR004087"
                     /db_xref="InterPro:IPR004088"
                     /db_xref="UniProt/TrEMBL:Q5NW02"
                     /translation="MQEHIKVPQDRIGAIIGVDGKIKELLEKKSGAQISVDSESGTVV
                     IDSKEGDPFKALKASDAIKAIARGFSPEKALKLLDSEDLILDMMDLSKITDTPSDLTR
                     IKGRIIGRGGKTREIIESMTGAKISVYGKTISIIGDAEQIMTVRTALDMLIDGAPHGA
                     VYGYLERRRREIKQSQFESIQ"
     gene            complement(10352. .10579)
                     /locus_tag="orf5"
     CDS             complement(10352. .10579)
                     /locus_tag="orf5"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAH04805.1"
                     /db_xref="GI:56295563"
                     /db_xref="UniProt/TrEMBL:Q5NW01"
                     /translation="MTTVRKMNVTSKNHTIATKNSTISILISMLTPLFNTTLSLLCHM
                     VFAWARRVITNIYAISEKCRYFRHIYRYNPL"
     gene            complement(10753. .11517)
                     /locus_tag="orf6"
     CDS             complement(10753. .11517)
                     /locus_tag="orf6"
                     /EC_number="4.2.1.33"
                     /codon_start=1
                     /transl_table=11
                     /product="3-isopropylmalate dehydratase"
                     /protein_id="CAH04806.1"
                     /db_xref="GI:56295564"
                     /db_xref="GOA:Q5NW00"
                     /db_xref="UniProt/TrEMBL:Q5NW00"
                     /translation="MIETTVVEYREIETESPVMIVGLPGIGLVGKLVADQLIIELKAE
                     KFMEIFSPCLPPQVNVNPDGTIKLVSNELYFYKGNGQTPDMIFLVGDHQSVTNEGHYE
                     LAGIYLDIAERYKVRRIYTMGGYGTGKLIETPHVLYATNKPEIIEEIKSYGAIFNEGE
                     LSGGIIGASGLILGLGELRGMECVCFMGETSGYLVDPKSAQVVLAVLCKALNIEIDTK
                     ELEARAQEIEKIITKIKEMEQSQIPKDYPAEEKYIG"
     gene            complement(11691. .12494)
                     /gene="eIF2a"
                     /locus_tag="orf7"
     CDS             complement(11691. .12494)
                     /gene="eIF2a"
                     /locus_tag="orf7"
                     /note="COG1093 (SUI2)"
                     /codon_start=1
                     /transl_table=11
                     /product="translation initiation factor 2, subunit alpha"
                     /protein_id="CAH04807.1"
                     /db_xref="GI:56295565"
                     /db_xref="GOA:Q5NVZ9"
                     /db_xref="UniProt/TrEMBL:Q5NVZ9"
                     /translation="MVKGWPNPGELVVCTVTKVVDFGAFVALDEYDNKEGLIHISEVA
                     SGWVKYIRDHVREGQKIVCKVLDVNPKRGHIDLSFKDVNEHQRRETIQLWKNDQKALK
                     WLQFVAEETKLQPKELEDLEDTIRDEFGNLYAVFEEVVHDGPQGLIKAGIPQNIADAI
                     NRIAHENVKVPSVDIAGYVDLLCPTSNGVEIIKKALKAAGSIDAGEGVTVDVTYVGAP
                     RYRIHVIAPDYKKAEKVLNTAARAAIDYVKKHEGEGEFHRHMEEAKSIS"
     gene            complement(12688. .12969)
                     /gene="RPL42A"
                     /locus_tag="orf8"

  Query Match            52.7%;  Score 31.6;  DB 3;  Length 36858;
  Best Local Similarity  74.1%;  Pred. No. 4.2;
  Matches   40;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||| ||||| ||||| | ||| || ||   ||| ||||| ||||| | ||| ||
Db      21891 GAACGGGAATCCGAACCCGAACCCTCTGAAGAACGGGAATCCGAACCCGAATCC 21838
```

```
RESULT 9
AC074138/c
LOCUS       AC074138              200418 bp    DNA     linear   PRI 30-NOV-2001
DEFINITION  Homo sapiens chromosome 19 clone CTD-3234P18, complete sequence.
ACCESSION   AC074138
VERSION     AC074138.5  GI:17155035
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 200418)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200418)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 200418)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 30, 2001 this sequence version replaced gi:15920100.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .200418
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTD-3234P18"
ORIGIN

  Query Match            52.7%;  Score 31.6;  DB 8;  Length 200418;
  Best Local Similarity  74.1%;  Pred. No. 4.8;
  Matches   40;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              |||||||||||  ||||  |||       |||||||||||| ||||  ||||
Db      67245 GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAACCGAAAGGGAAG 67192


RESULT 10
AR503827/c
LOCUS       AR503827                466 bp    DNA     linear   PAT 22-SEP-2004
DEFINITION  Sequence 8787 from patent US 6703491.
ACCESSION   AR503827
VERSION     AR503827.1  GI:52439302
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L.,
            Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.
  TITLE     Drosophila sequences
  JOURNAL   Patent: US 6703491-A 8787 09-MAR-2004;
            Exelixis, Inc.; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

```
  Query Match             51.7%;  Score 31;  DB 6;  Length 466;
  Best Local Similarity   72.7%;  Pred. No. 4.9;
  Matches   40;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      |||||  | ||||  |||| |||| |
Db        189 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 135


RESULT 11
AR519109/c
LOCUS       AR519109                466 bp    DNA     linear   PAT 22-SEP-2004
DEFINITION  Sequence 24069 from patent US 6703491.
ACCESSION   AR519109
VERSION     AR519109.1  GI:52454584
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L.,
            Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.
  TITLE     Drosophila sequences
  JOURNAL   Patent: US 6703491-A 24069 09-MAR-2004;
            Exelixis, Inc.; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             51.7%;  Score 31;  DB 6;  Length 466;
  Best Local Similarity   72.7%;  Pred. No. 4.9;
  Matches   40;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      |||||  | ||||  |||| |||| |
Db        189 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 135


RESULT 12
AC014250/c
LOCUS       AC014250               26002 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC014250
VERSION     AC014250.1  GI:6437085
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 26002)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214205 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .26002
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

  Query Match             51.7%;  Score 31;  DB 14;  Length 26002;
  Best Local Similarity   72.7%;  Pred. No. 6.6;
  Matches   40;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      |||||  | ||||  |||| |||| |
Db      25074 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 25020


RESULT 13
AC010581
LOCUS       AC010581               175053 bp    DNA     linear   INV 27-APR-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 85D-85D, BAC clone
            BACR19J06, complete sequence.
ACCESSION   AC010581
VERSION     AC010581.3  GI:13811835
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 175053)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 85D-85D
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175053)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Apr 27, 2001 this sequence version replaced gi:6957905.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .175053
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="85D-85D"
                     /clone="BACR19J06 (D834)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
ORIGIN
```

```
  Query Match             51.7%;  Score 31;  DB 2;  Length 175053;
  Best Local Similarity   72.7%;  Pred. No. 7.6;
  Matches   40;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      ||||| | ||||  |||| |||| |
Db     100039 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 100093


RESULT 14
AE003683
LOCUS       AE003683               243348 bp    DNA     linear   INV 05-AUG-2005
DEFINITION  Drosophila melanogaster chromosome 3R, section 21 of 118 of the
            complete sequence.
ACCESSION   AE003683 AE002708 AE014297
VERSION     AE003683.3  GI:23170805
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 243348)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,
            Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,
            Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,
            Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S.,
            Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P.,
            Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A.,
            Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,
            Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,
            Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S.,
            Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
            Ferriera,S., Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., WoodageT,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
   PUBMED   10731132
REFERENCE   2  (bases 1 to 243348)
  AUTHORS   Celniker,S.E., Wheeler,D.A., Kronmiller,B., Carlson,J.W.,
            Halpern,A., Patel,S., Adams,M., Champe,M., Dugan,S.P., Frise,E.,
            Hodgson,A., George,R.A., Hoskins,R.A., Laverty,T., Muzny,D.M.,
            Nelson,C.R., Pacleb,J.M., Park,S., Pfeiffer,B.D., Richards,S.,
            Sodergren,E.J., Svirskas,R., Tabor,P.E., Wan,K., Stapleton,M.,
            Sutton,G.G., Venter,C., Weinstock,G., Scherer,S.E., Myers,E.W.,
            Gibbs,R.A. and Rubin,G.M.
  TITLE     Finishing a whole-genome shotgun: release 3 of the Drosophila
            melanogaster euchromatic genome sequence
  JOURNAL   Genome Biol. 3 (12), RESEARCH0079 (2002)
   PUBMED   12537568
REFERENCE   3  (bases 1 to 243348)
  AUTHORS   Misra,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S.,
            Hradecky,P., Huang,Y., Kaminker,J.S., Millburn,G.H., Prochnik,S.E.,
            Smith,C.D., Tupy,J.L., Whitfied,E.J., Bayraktaroglu,L.,
            Berman,B.P., Bettencourt,B.R., Celniker,S.E., de Grey,A.D.,
            Drysdale,R.A., Harris,N.L., Richter,J., Russo,S., Schroeder,A.J.,
            Shu,S.Q., Stapleton,M., Yamada,C., Ashburner,M., Gelbart,W.M.,
            Rubin,G.M. and Lewis,S.E.
  TITLE     Annotation of the Drosophila melanogaster euchromatic genome: a
            systematic review
  JOURNAL   Genome Biol. 3 (12), RESEARCH0083 (2002)
   PUBMED   12537572
REFERENCE   4  (bases 1 to 243348)
  AUTHORS   Kaminker,J.S., Bergman,C.M., Kronmiller,B., Carlson,J.,
            Svirskas,R., Patel,S., Frise,E., Wheeler,D.A., Lewis,S.E.,
            Rubin,G.M., Ashburner,M. and Celniker,S.E.
  TITLE     The transposable elements of the Drosophila melanogaster
            euchromatin: a genomics perspective
  JOURNAL   Genome Biol. 3 (12), RESEARCH0084 (2002)
   PUBMED   12537573
REFERENCE   5  (bases 1 to 243348)
  AUTHORS   Celniker,S., Carlson,J., Wan,K., Pfeiffer,B., Frise,E., George,R.,
            Hoskins,R., Stapleton,M., Pacleb,J., Park,S., Svirskas,R.,
            Smith,E., Yu,C. and Rubin,G.
  CONSRTM   Berkeley Drosophila Genome Project
  TITLE     Drosophila melanogaster release 4 sequence
  JOURNAL   Unpublished
REFERENCE   6  (bases 1 to 243348)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
REFERENCE   7  (bases 1 to 243348)
  AUTHORS   .
  CONSRTM   FlyBase
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-2005) FlyBase, Harvard University, Biological
            Laboratories, 16 Divinity Ave, Cambridge, MA 02138, USA
COMMENT     On Sep 18, 2002 this sequence version replaced gi:10726407.
FEATURES             Location/Qualifiers
     source          1. .243348
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
     gene            complement(<4089. .>13382)
                     /gene="CG16779"
                     /locus_tag="Dmel_CG16779"
                     /old_locus_tag="CG16779"
                     /map="85D17-85D17"
                     /db_xref="FLYBASE:FBgn0037698"
     mRNA            complement(join(<4089. .4230,5579. .6003,6228. .6386,
                     6450. .9500,9567. .9938,10004. .10763,10885. .11003,
                     12316. .12980,13133. .>13382))
                     /gene="CG16779"
                     /locus_tag="Dmel_CG16779"
                     /old_locus_tag="CG16779"
                     /product="CG16779-RA"
                     /db_xref="FLYBASE:FBgn0037698"
     CDS             complement(join(4089. .4230,5579. .6003,6228. .6386,
                     6450. .9500,9567. .9938,10004. .10763,10885. .11003,
                     12316. .12980,13133. .13382))
                     /gene="CG16779"
                     /locus_tag="Dmel_CG16779"
                     /old_locus_tag="CG16779"
                     /note="CG16779 gene product from transcript CG16779-RA"
                     /codon_start=1
                     /product="CG16779-PA"
                     /protein_id="AAF54383.1"
                     /db_xref="GI:7299186"
                     /db_xref="FLYBASE:FBgn0037698"
```

```
                     /translation="MCSASVGTEGSVTLQLRDANALEAAATKTATAAAAAAAAAAATS
                     TTSPNGVAGAPASTLENALTIGYPDPEMLADVLGTIQTASLKNNNSITAATSSNSSTK
                     STSSNHLTNPNKITIKRRSVHSVSTSTNSSSISSCSSSSSSSSNSGKSNTSYIKNCLIR
                     SSSCSNTVTNVTTLAQIMSNSSTSSAASLLNQHNNNSNCSNSSNFSTASSVGSSISVG
                     SNSSSSNCSNGSNSNSNDSNSSSGGGHRLSFKGTGRHVPASGLSSSAASTNTNGTVSG
                     IGGIISVEAPRKNRPKLSSPTRHGPQQCQICCKIFGNASALAKHKLTHSDERKYICAL
                     CSKAFKRQDHLNGHMMTHRNKKPYECKADGCGKSYCDARSLRRHSENHHAGGAATPTT
                     SQSLSPTASSSGTSNSSGAGAATSSLSLSPATASGDASSPDGATCIRTYISTGSSVVD
                     AATGIALSDEQIKAMNLPIKTGVTLLSPTTSTSSNASSSASSASSLSATGSLPGSGS
                     GSGSVSMAPSPSSGSVITACSPTITLSDGMSLEGEGLTREQLDLISKIMQQTKQTSAQ
                     VTVSSPTSVSSYKINTNSSPSQNRPRTWNMQLLNSAQNVTVTVEDNSELVTPSSNSPV
                     EVKEEELSSQLVGTMNPHLLNIVKLDKPVECNLCHRKFKNIPALNGHMRLHGGYYKKD
                     PETKRSEKKDSSGPPLQTASIGVRALIEEKIISKRKDMTKGSFVVPAPPHSSGTTTTL
                     RRSISDLESPLNPKTSSSSHTQTMTTTTGTTTAVLPAATTIKSSNGLSIQQIGLPQSI
                     EIFSGGQKQAKTLNLGSGTNTITITTNNVPTTTTMSALTALKAGGTISGISTATNTDP
                     KDSTLIELLKRGTRIAVTSKKNQSHGQTGSSTMMSSSGAATNAVGAVTELTTIGAGGG
                     VQTVGRQIITNNNRTVIIPSDVQVVSTKSKLSSLSSLVKCNMTNSGTSSSGNLSLADG
                     TPLSLTIASNQELTGGGGNSSGSGSVITSGGGGVYTVTYTSDGTDLFDDAEVYNVSDT
                     EMLLQTVDSMELLNDEEIKSEHSEDFALLSEAGDSAHTQLVKLEPESGQANSGSGSVS
                     GANTTPLPTFQQFHSKELIMQNSSQIQAIASMRGSGGGVLASPLHSPLAYPTPPSSHE
                     NMAQSSPFIEDAAAQFVDASNTFFGDKTDFSHIYFKTDEGEATIEQLNEHDNEKILKL
                     KSVLEDSSFDPSIKVEDLLNGTDDDTECDLREFAETNLSFLDEDQEFLNDSRNATSPL
                     SESFFTSGIGSAEDVKQVLREVLPDENMQLQLSSEQQGENIIDLYYLPGLGLQSQMMP
                     NSEDPLLSSSPREFGQQRQQLPMQTTNMQQPMEQTLQDNNNIYQQQEQQQEQHQQQQS
                     QQEQPPQQEQQQQHHQQLLMPQQPLNQTESLPAVGQQQGDFMLPLVGGHGFTQVTSQT
                     QYIDNSQGGMGMQPLNSLLQPLLYGGATTSNGNASAGGNESLLTTDCQMNANQGQIDT
                     GLMFACGNTTTMANKSLAAIVSNPPSVANLQPLSNQTNSILKRRLRSNAPQETHKFSK
                     FHTLSPHRSKLRKPSRTHYTPAPILNPDRKGTGLYCNVRKQLGQGLFDAFDDDFGDSV
                     GLVDFSDESKVNLGSTYQAQIPSCRPLEEALRDSPGAEMMWNPEVQEDEKILMRYIDL
                     SKSSAVPMGSHSEEVALQTLLRAKGNSAAAVLSLLQTQSGAFQMKWTAYELEQFLRGL
                     EKNGKDFGKIASELQTKSSGECVQMYYFWKKLCVDYKVTHLKMEPVVVIAPAVEKPYV
                     CEIADCSASFSSKAALHGHVRIHAYGRSASSNTTNNNSSSSNTGSSSNVQHAAASSAS
                     SGGNNNNSSSYACATLSGSNGSSNNASTNVAGAVATHSSNMAQNGSNCSGHANAAIAT
                     SALSTTPKLEAGMGNTTNIPIAKEGEFPCKVCGKVFNKVKSRSAHMKTHRVQEPDQKH
                     QLQNQAGAGLTTVATSTITAPS"
     repeat_region   complement(12508. .13312)
                     /transposon="TART-Dvir-like{}2363"
                     /db_xref="FLYBASE:FBti0060424"
     gene            32402. .34329
                     /gene="CG8147"
                     /locus_tag="Dmel_CG8147"
                     /old_locus_tag="CG8147"
                     /map="85D18-85D18"
                     /db_xref="FLYBASE:FBgn0043791"
     mRNA            join(32402. .32545,32603. .34329)
                     /gene="CG8147"
                     /locus_tag="Dmel_CG8147"
                     /old_locus_tag="CG8147"
                     /product="CG8147-RA"
                     /db_xref="FLYBASE:FBgn0043791"
     CDS             32629. .34188
                     /gene="CG8147"
                     /locus_tag="Dmel_CG8147"
                     /old_locus_tag="CG8147"
                     /note="CG8147 gene product from transcript CG8147-RA"
                     /codon_start=1
                     /product="CG8147-PA"
                     /protein_id="AAF54384.3"
                     /db_xref="GI:28381202"
                     /db_xref="FLYBASE:FBgn0043791"
                     /translation="MHPVFSLMGTGSNPIGGGRYKRAMPQIVFDAVRSEERYAEYWQG
                     LAAQTLDQQLESKLRLNTQLARNVMLFIGDGMSIPTITAGRVYLGGEEKQFAFEQFPY
                     VGLSKTYCANMQVADSACTATAYLGGVKANYGTIGVSAAVQFKDCQAQAQAAHHVSSI
                     AAWAQKQGMATGLVTTTSVTHASPAGVYAHLANRNWENDAEVVGDNGDPDLCPDAAAQ
                     LINSPVGQKLNVIMGGGRENFLPKGVTDSSGAPGRRLDGRNLIDEWKNQHTNSAQYVE

  Query Match             51.7%;  Score 31;  DB 2;  Length 243348;
  Best Local Similarity   72.7%;  Pred. No. 7.8;
  Matches   40;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      ||||| | ||||  |||| |||| |
Db       1983 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 2037
```

```
RESULT 15
AC079986/c
LOCUS       AC079986              180889 bp    DNA     linear   HTG 04-FEB-2001
DEFINITION  Homo sapiens chromosome RPCI-11 clone RP11-600E3, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC079986
VERSION     AC079986.2  GI:12225354
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 180889)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180889)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 14, 2001 this sequence version replaced gi:10198579.

            ---------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0600E03
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 175434 bases at least Q40
            Consensus quality: 177921 bases at least Q30
            Consensus quality: 179200 bases at least Q20
            Insert size: 187000; agarose-fp
            Insert size: 180389; sum-of-contigs
            Quality coverage: 5.44 in Q20 bases; agarose-fp
            Quality coverage: 5.71 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1895: contig of 1895 bp in length
            *     1896     1995: gap of unknown length
            *     1996     4214: contig of 2219 bp in length
            *     4215     4314: gap of unknown length
            *     4315     8326: contig of 4012 bp in length
            *     8327     8426: gap of unknown length
            *     8427    25575: contig of 17149 bp in length
            *    25576    25675: gap of unknown length
            *    25676    60037: contig of 34362 bp in length
            *    60038    60137: gap of unknown length
            *    60138   180889: contig of 120752 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180889
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="RPCI-11"
                     /clone="RP11-600E3"
```

```
     misc_feature    1. .1895
                     /note="assembly_name:Contig12"
     gap             1896. .1995
                     /estimated_length=unknown
     misc_feature    1996. .4214
                     /note="assembly_name:Contig13
                     clone_end:T7
                     vector_side:left"
     gap             4215. .4314
                     /estimated_length=unknown
     misc_feature    4315. .8326
                     /note="assembly_name:Contig14"
     gap             8327. .8426
                     /estimated_length=unknown
     misc_feature    8427. .25575
                     /note="assembly_name:Contig15"
     gap             25576. .25675
                     /estimated_length=unknown
     misc_feature    25676. .60037
                     /note="assembly_name:Contig16"
     gap             60038. .60137
                     /estimated_length=unknown
     misc_feature    60138. .180889
                     /note="assembly_name:Contig17
                     clone_end:SP6
                     vector_side:left"
ORIGIN

  Query Match             51.0%;  Score 30.6;  DB 14;  Length 180889;
  Best Local Similarity   73.6%;  Pred. No. 10;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              |||||||||||||| ||||  |||       | ||||||||||  |||| ||||
Db     139706 GGGAAAGGGAAACGGAAATGGAAAGTGACAGAGAAAGGGAAAGGGAAAGTGAA 139654


Search completed: March 30, 2006, 17:41:29
Job time : 1732 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2006, 15:25:47 ; Search time 1170 Seconds
(without alignments)
341.779 Million cell updates/sec

Title: US-10-677-777-5
Perfect score: 60
Sequence: 1 ctcgggaaagggaaaccgaa.....ggaaaccgaaactgaagccc 60

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	60	100.0	60	12	ADO43723	Ado43723 Nucleotid
	2	60	100.0	4412	12	ADO43721	Ado43721 Nucleotid
	3	60	100.0	5849	12	ADO43719	Ado43719 Nucleotid
c	4	30.6	51.0	1413	6	ABQ32828	Abq32828 Oligonucl
	5	30.6	51.0	1413	6	ABQ32829	Abq32829 Oligonucl
	6	30.6	51.0	1418	6	ABQ32843	Abq32843 Oligonucl
c	7	30.6	51.0	1418	6	ABQ32842	Abq32842 Oligonucl
	8	30	50.0	33	2	AAQ26486	Aaq26486 Oligonucl
	9	30	50.0	33	3	AAZ90328	Aaz90328 Electroph
	10	30	50.0	34	2	AAX15167	Aax15167 Oligonucl
	11	30	50.0	40	2	AAQ89791	Aaq89791 ISRE of I
c	12	30	50.0	2308	6	ABQ39166	Abq39166 Oligonucl
	13	30	50.0	2308	6	ABQ39167	Abq39167 Oligonucl
c	14	30	50.0	2775	12	ADO36037	Ado36037 Novel mou
c	15	30	50.0	6061	6	ABL33564	Abl33564 Human imm
c	16	30	50.0	6061	6	ABL70593	Abl70593 Chemicall
c	17	30	50.0	6061	6	AAS61310	Aas61310 Human gen
c	18	30	50.0	6510	4	AAS46647	Aas46647 Tumour su
c	19	30	50.0	6510	6	ABL32394	Abl32394 Human imm
c	20	30	50.0	6510	6	AAS61114	Aas61114 Human gen
c	21	30	50.0	69515	11	ACN44660	Acn44660 Mouse gen
	22	30	50.0	110000	10	ADG70447_1	Continuation (2 of
	23	30	50.0	110000	10	ABZ79565_1	Continuation (2 of
	24	29	48.3	1067	6	ABQ36077	Abq36077 Oligonucl
c	25	29	48.3	1067	6	ABQ36076	Abq36076 Oligonucl
c	26	29	48.3	2365	6	ABQ17068	Abq17068 Oligonucl
	27	29	48.3	2365	6	ABQ17069	Abq17069 Oligonucl
c	28	29	48.3	2395	6	ABQ13720	Abq13720 Oligonucl
	29	29	48.3	2395	6	ABQ13721	Abq13721 Oligonucl
c	30	29	48.3	34460	11	ACN43924	Acn43924 Mouse gen
	31	29	48.3	110459	14	AEB44581	Aeb44581 Sleep dis
	32	29	48.3	114596	14	ADZ70593	Adz70593 Human cDN
	33	28.8	48.0	13497	4	ABL03557	Abl03557 Drosophil
c	34	28.8	48.0	16489	4	ABL03556	Abl03556 Drosophil
	35	28.4	47.3	543	3	AAA43729	Aaa43729 Mouse sec
	36	28.4	47.3	642	6	ABQ53867	Abq53867 Oligonucl
c	37	28.4	47.3	642	6	ABQ53866	Abq53866 Oligonucl
	38	28.4	47.3	688	10	ADD34808	Add34808 Mouse mit
	39	28.4	47.3	752	6	ABQ41389	Abq41389 Oligonucl
c	40	28.4	47.3	752	6	ABQ41388	Abq41388 Oligonucl
	41	28.2	47.0	311	4	AAI80200	Aai80200 Human pol
	42	28.2	47.0	4907	4	ABL06441	Abl06441 Drosophil
	43	28	46.7	30	2	AAX90996	Aax90996 ISRE regi
	44	28	46.7	30	3	AAZ92043	Aaz92043 STAT5 bin
	45	28	46.7	30	6	AAD44996	Aad44996 ISG15 oli

ALIGNMENTS

RESULT 1
ADO43723
ID ADO43723 standard; DNA; 60 BP.
XX
AC ADO43723;
XX
DT 29-JUL-2004 (first entry)
XX
DE Nucleotide sequence of an ISRE from human IGS15 gene.
XX
KW signal transduction; cell surface protein; ISRE-luc vector;
KW interferon stimulatory response element; ISG15 gene; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 41. .55
FT /*tag= a
FT /note= "consensus sequence"
XX
PN WO2004039990-A2.
XX
PD 13-MAY-2004.
XX
PF 03-OCT-2003; 2003WO-US031310.
XX
PR 04-OCT-2002; 2002US-0415818P.
XX
PA (NEUT-) NEUTEKBIO LTD.
PA (MCIN/) MCINNIS P.
XX
PI Tovey MG, Lallemand C;
XX
DR WPI; 2004-376199/35.
XX
PT Cell transformed with reporter gene construct for determining presence of
PT molecule that activates signal transduction activity, comprises
PT nucleotide sequence encoding reporter gene product linked to
PT transcriptional control elements.
XX
PS Disclosure; Page 90; 91pp; English.

```
XX
CC   The specification describes a cell transformed with a reporter gene
CC   construct, comprising a nucleotide sequence encoding a reporter gene
CC   product operatively linked to one or more transcriptional control
CC   elements that is regulated by the signal transduction activity of a cell
CC   surface protein in response to an extracellular signal. The cell is
CC   treated so that it will maintain the signal transduction activity for at
CC   least 1 hour but not more than 30 days. The cells of the invention are
CC   useful for determining the presence or the level in a sample of a
CC   molecule that activates the signal transduction activity of a cell
CC   surface protein. The present represents an interferon stimulatory
CC   response element (ISRE) from the human ISG15 gene. The ISRE is used to
CC   produce cells of the invention.
XX
SQ   Sequence 60 BP; 22 A; 16 C; 18 G; 4 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 60;  DB 12;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-11;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60


RESULT 2
ADO43721
ID   ADO43721 standard; DNA; 4412 BP.
XX
AC   ADO43721;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Nucleotide sequence of ISRE-EGFP vector.
XX
KW   signal transduction; cell surface protein; ISRE-luc vector;
KW   enhanced green fluorescent protein; gene;
KW   interferon stimulatory response element; ISG15 gene; ss; EGFP.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    38. .89
FT                   /*tag=  a
FT                   /note= "interferon stimulatory response element from the
FT                   ISG15 gene"
FT   promoter        95. .290
FT                   /*tag=  b
FT                   /note= "SV40 minimal promoter"
FT   CDS             358. .1077
FT                   /*tag=  c
FT                   /note= "EGFP"
XX
PN   WO2004039990-A2.
XX
PD   13-MAY-2004.
XX
PF   03-OCT-2003; 2003WO-US031310.
XX
PR   04-OCT-2002; 2002US-0415818P.
XX
PA   (NEUT-) NEUTEKBIO LTD.
PA   (MCIN/) MCINNIS P.
XX
PI   Tovey MG,  Lallemand C;
XX
DR   WPI; 2004-376199/35.
DR   P-PSDB; ADO43722.
XX
PT   Cell transformed with reporter gene construct for determining presence of
PT   molecule that activates signal transduction activity, comprises
PT   nucleotide sequence encoding reporter gene product linked to
PT   transcriptional control elements.
XX
PS   Example; Page 87-89; 91pp; English.
XX
CC   The specification describes a cell transformed with a reporter gene
CC   construct, comprising a nucleotide sequence encoding a reporter gene
CC   product operatively linked to one or more transcriptional control
CC   elements that is regulated by the signal transduction activity of a cell
CC   surface protein in response to an extracellular signal. The cell is
CC   treated so that it will maintain the signal transduction activity for at
CC   least 1 hour but not more than 30 days. The cells of the invention are
CC   useful for determining the presence or the level in a sample of a
CC   molecule that activates the signal transduction activity of a cell
CC   surface protein. The present represents the ISRE-EGFP vector, comprising
CC   an enhanced green fluorescent protein (EGFP) gene reporter construct and
CC   an interferon stimulatory response element (ISRE) from the ISG15 gene.
CC   This vector is used to produce cells of the invention.
XX
SQ   Sequence 4412 BP; 1055 A; 1225 C; 1165 G; 967 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 60;  DB 12;  Length 4412;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-10;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         30 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 89


RESULT 3
ADO43719
ID   ADO43719 standard; DNA; 5849 BP.
XX
AC   ADO43719;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Nucleotide sequence of ISRE-luc vector.
XX
KW   signal transduction; cell surface protein; ISRE-luc vector; luciferase;
KW   gene; interferon stimulatory response element; ISG15 gene; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    38. .97
FT                   /*tag=  a
FT                   /note= "interferon stimulatory response element from the
FT                   ISG15 gene"
FT   promoter        103. .288
FT                   /*tag=  b
FT                   /note= "SV40 minimal promoter"
FT   CDS             328. .1980
FT                   /*tag=  c
FT                   /product= "luciferase"
XX
PN   WO2004039990-A2.
XX
PD   13-MAY-2004.
XX
PF   03-OCT-2003; 2003WO-US031310.
XX
PR   04-OCT-2002; 2002US-0415818P.
XX
PA   (NEUT-) NEUTEKBIO LTD.
PA   (MCIN/) MCINNIS P.
XX
PI   Tovey MG,  Lallemand C;
XX
DR   WPI; 2004-376199/35.
DR   P-PSDB; ADO43720.
XX
PT   Cell transformed with reporter gene construct for determining presence of
PT   molecule that activates signal transduction activity, comprises
```

PT nucleotide sequence encoding reporter gene product linked to
PT transcriptional control elements.
XX
PS Example; Page 80-84; 91pp; English.
XX
CC The specification describes a cell transformed with a reporter gene
CC construct, comprising a nucleotide sequence encoding a reporter gene
CC product operatively linked to one or more transcriptional control
CC elements that is regulated by the signal transduction activity of a cell
CC surface protein in response to an extracellular signal. The cell is
CC treated so that it will maintain the signal transduction activity for at
CC least 1 hour but not more than 30 days. The cells of the invention are
CC useful for determining the presence or the level in a sample of a
CC molecule that activates the signal transduction activity of a cell
CC surface protein. The present represents the ISRE-luc vector, comprising a
CC luciferase gene reporter construct and an interferon stimulatory response
CC element (ISRE) from the ISG15 gene. This vector is used to produce cells
CC of the invention.
XX
SQ Sequence 5849 BP; 1581 A; 1343 C; 1327 G; 1598 T; 0 U; 0 Other;

Query Match 100.0%; Score 60; DB 12; Length 5849;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
||||||

CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1413 BP; 594 A; 474 C; 172 G; 173 T; 0 U; 0 Other;

Query Match 51.0%; Score 30.6; DB 6; Length 1413;
Best Local Similarity 73.6%; Pred. No. 1.1;
Matches 39; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
|||| |||||||||||| ||| || |||| |||||||||||| ||| |
Db 253 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 305

RESULT 6
ABQ32843
ID ABQ32843 standard; DNA; 1418 BP.
XX
AC ABQ32843;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19434.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1418 BP; 589 A; 487 C; 173 G; 169 T; 0 U; 0 Other;

Query Match 51.0%; Score 30.6; DB 6; Length 1418;
Best Local Similarity 73.6%; Pred. No. 1.1;
Matches 39; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
|||| |||||||||||| ||| || |||| |||||||||||| ||| |
Db 344 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 396

RESULT 7
ABQ32842/c
ID ABQ32842 standard; DNA; 1418 BP.
XX
AC ABQ32842;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19433.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms

```
CC   (SNP's); and (ii) for differentiation of cell or tissue types and for
CC   investigating cell differentiation. The method allows the methylation
CC   status of many C residues to be determined simultaneously. ABQ13410-
CC   ABQ54121 represent genomic DNA sequences used to illustrate the method
CC   for determining the degree of cytosine methylation described in the
CC   disclosure of the invention
XX
SQ   Sequence 1418 BP; 169 A; 173 C; 487 G; 589 T; 0 U; 0 Other;

  Query Match           51.0%;  Score 30.6;  DB 6;  Length 1418;
  Best Local Similarity  73.6%;  Pred. No. 1.1;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  |||||||||||| ||| ||      ||||  |||||||||||| ||| |
Db       1075 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 1023


RESULT 8
AAQ26486
ID   AAQ26486 standard; DNA; 33 BP.
XX
AC   AAQ26486;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   14-JAN-1993  (first entry)
XX
DE   Oligonucleotide .015 for HNF-4 mobility shift assay.
XX
KW   hepatocyte nuclear factor 4; apolipoproteins; fat; cholesterol; Apo CIII;
KW   Apo A1; Apo B; pyruvate kinase; alpha 1 antitrypsin;
KW   glutamine synthetase; coronary heart disease hyperlipidaemia;
KW   liver disease; arteriosclerosis; obesity; ss.
XX
OS   Rattus sp.
XX
PN   WO9211365-A1.
XX
PD   09-JUL-1992.
XX
PF   23-DEC-1991;   91WO-US009733.
XX
PR   21-DEC-1990;   90US-00631720.
XX
PA   (UYRQ ) UNIV ROCKEFELLER.
XX
PI   Sladek FM,  Zhong W,  Darnell JE;
XX
DR   WPI; 1992-250087/30.
XX
PT   Hepatocyte nuclear factor 4 and its DNA, regulation and antibodies -
PT   useful for treating cardiovascular diseases e.g. atherosclerotic heart
PT   disease, hyperlipidaemia and arteriosclerosis; also as an anorectic.
XX
PS   Disclosure; Page 21; 100pp; English.
XX
CC   This sequence was used in a mobility shift assay using purified HNF-4 as
CC   an unlabelled competitor probe. HNF-4 was found not to significantly bind
CC   the oestrogen, thyroid hormone and glucocorticoid response element.
CC   (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-MAR-2003
CC   to correct PN field.)
XX
SQ   Sequence 33 BP; 12 A; 8 C; 10 G; 3 T; 0 U; 0 Other;

  Query Match           50.0%;  Score 30;  DB 2;  Length 33;
  Best Local Similarity  100.0%;  Pred. No. 0.91;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 9
AAZ90328
ID   AAZ90328 standard; DNA; 33 BP.
XX
AC   AAZ90328;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Electrophoretic shift mobility assay negative control probe.
XX
KW   Hepatocyte nuclear factor 4; HNF-4; rat; transcription factor;
KW   ligand binding domain; transthyretin; apolipoprotein CIII;
KW   apoCIII expression; agonist; antagonist; very low density lipoprotein;
KW   VLDL; cholesterol; cardiovascular disease; liver disease; obesity;
KW   electrophoretic mobility shift assay; EMSA; probe; ds.
XX
OS   Synthetic.
XX
PN   US6025196-A.
XX
PD   15-FEB-2000.
XX
PF   11-MAR-1998;   98US-00038217.
XX
PR   23-DEC-1991;   91WO-US009733.
PR   28-OCT-1993;   93US-00078222.
PR   14-JUN-1996;   96US-00661330.
XX
PA   (UYRQ ) UNIV ROCKEFELLER.
XX
PI   Darnell JE,  Zhong W,  Sladek FM;
XX
DR   WPI; 2000-181818/16.
XX
PT   Chimeric construct expresses a polypeptide comprising the ligand binding
PT   domain of a rat hepatocyte nuclear factor 4 receptor and a DNA binding
PT   domain of another receptor.
XX
PS   Disclosure; Col 10; 48pp; English.
XX
CC   The invention relates to a novel chimeric construct which expresses a
CC   protein comprising the ligand binding domain of a rat hepatocyte nuclear
CC   factor 4 (HNF-4) receptor and a DNA binding domain of another receptor
CC   which has a known ligand. Rat HNF-4 binds to sites required for the
CC   transcription of the transthyretin and apolipoprotein CIII (apoCIII)
CC   genes. ApoCIII is a constituent of very low density lipoprotein (VLDL)
CC   and high density lipoprotein (HDL) and is a noncompetitive inhibitor of
CC   the activity of lipoprotein lipase. Agonists and antagonists acting on
CC   the rat hepatocyte nuclear factor 4 (HNF-4) interfere with the deposition
CC   of VLDL and cholesterol in blood vessels. The chimeric construct is
CC   useful for expressing the ligand binding domain of rat HNF-4. Agonists
CC   and antagonists of rat HNF-4 are useful in the treatment of
CC   cardiovascular disease, such as atherosclerotic heart disease and
CC   arteriosclerosis, liver disease (involving the presence of excessive
CC   lipid levels) and obesity. Sequences AAZ90328-Z90337 represent
CC   electrophoretic mobility shift assay probes used to characterise purified
CC   rat liver HNF-4 protein in the exemplifications of the present invention
XX
SQ   Sequence 33 BP; 12 A; 8 C; 10 G; 3 T; 0 U; 0 Other;

  Query Match           50.0%;  Score 30;  DB 3;  Length 33;
  Best Local Similarity  100.0%;  Pred. No. 0.91;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 10
AAX15167
ID   AAX15167 standard; DNA; 34 BP.
```

XX
AC AAX15167;
XX
DT 22-APR-1999 (first entry)
XX
DE Oligonucleotide ISRE-1A used in binding studies with LDL.
XX
KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW non-small cell lung carcinoma; diabetes; arteriosclerosis; ss.
XX
OS Synthetic.
XX
PN WO9856938-A1.
XX
PD 17-DEC-1998.
XX
PF 10-JUN-1998; 98WO-US011927.
XX
PR 13-JUN-1997; 97US-00874807.
PR 14-MAY-1998; 98US-00079030.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Guevara JG, Hoogeveen RC, Moore JP;
XX
DR WPI; 1999-070331/06.
XX
PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT used for delivering nucleic acid to cells for gene therapy and antisense
PT treatment.
XX
PS Example 5; Page 75; 293pp; English.
XX
CC AAX15164-63 represent oligonucleotides used in binding studies with low
CC density lipoprotein (LDL), in the course of the invention. The
CC specification describes a composition that comprises low density
CC lipoprotein (LDL) and apolipoproteins for the binding and in vivo
CC transport of nucleic acids. The composition is used to deliver nucleic
CC acids to eukaryotic cells, in vivo or in vitro, for expressing a
CC therapeutic polypeptide or antisense molecule (or ribozyme). Specifically
CC they are used for gene therapy of cancers (particularly non-small cell
CC lung carcinoma), diabetes, cystic fibrosis and arteriosclerosis
XX
SQ Sequence 34 BP; 12 A; 8 C; 11 G; 3 T; 0 U; 0 Other;

Query Match 50.0%; Score 30; DB 2; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
||||||||||||||||||||||||||||||
Db 4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33

RESULT 11
AAQ89791
ID AAQ89791 standard; DNA; 40 BP.
XX
AC AAQ89791;
XX
DT 19-DEC-1995 (first entry)
XX
DE ISRE of ISG15.
XX
KW Cytokine; signal transducer and activator of transcription; STAT;
KW transcriptional factor family; receptor; phosphorylated tyrosine;
KW inteferon-gamma receptor alpha chain; allograft rejection; disorder;
KW insuling dependent diabetes mellitus; immune response; cancer; PCR;
KW inflammatory response; primer; amplification; probe; ss.
XX
OS Synthetic.
XX
PN WO9511312-A1.
XX
PD 27-APR-1995.
XX
PF 21-OCT-1994; 94WO-US012095.
XX
PR 22-OCT-1993; 93US-00141499.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Schreiber RD, Farrar MA, Greenlund AC;
XX
DR WPI; 1995-170231/22.
XX
PT Inhibiting cytokine-induced activation of a STAT family transcription
PT factor - by admin. of a peptide derived from the cytokine receptor and
PT contg. phospho:tyrosine, esp. for inhibiting induction by interferon-
PT gamma.
XX
PS Example 4; Page 29; 65pp; English.
XX
CC This oligonucleotide was used as a control in an electrophoretic mobility
CC shift assay (EMSA) to determine the biological role of inteferon(IFN)-
CC gamma induced-tyrosine phosphorylation at residue Tyr440 of the IFN-gamma
CC receptor using the gamma response region (GRR) of the Fc(gamma)RI gene as
CC a probe (see AAQ89789-90). Peptides based on the sequence around the
CC Tyr440 (see AAR73625-31) can be used to inhibit the IFN-gamma induced
CC activation of members of the cytokine inducible signal transducers and
CC activators of transcription (STAT) family of transcriptional factors. The
CC inhibitory peptides are useful in the treatment of e.g. insulin dependent
CC diabetes mellitus, blocking allograft rejection and in treating other
CC disorders involving an uncontrolled immune response or inflammatory
CC response. The peptides can also be used to treat cancer by being targeted
CC to the cancerous cells
XX
SQ Sequence 40 BP; 14 A; 11 C; 11 G; 4 T; 0 U; 0 Other;

Query Match 50.0%; Score 30; DB 2; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.95;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
||||||||||||||||||||||||||||||
Db 9 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 38

RESULT 12
ABQ39166/c
ID ABQ39166 standard; DNA; 2308 BP.
XX
AC ABQ39166;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25757.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.

```
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful for
PT   diagnosis and prognosis, comprises selective hybridization of amplicons
PT   from chemically treated DNA.
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one member,
CC   of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC   degree of hybridisation to both classes is determined from the label on
CC   the amplicon. From the ratio of labels hybridised to the two classes of
CC   oligomers, the degree of methylation is calculated. The method is used:
CC   (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC   and of a wide range of diseases, e.g. cancer, disorders of the central
CC   nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC   particularly by detecting mutations or single nucleotide polymorphisms
CC   (SNP's); and (ii) for differentiation of cell or tissue types and for
CC   investigating cell differentiation. The method allows the methylation
CC   status of many C residues to be determined simultaneously. ABQ13410-
CC   ABQ54121 represent genomic DNA sequences used to illustrate the method
CC   for determining the degree of cytosine methylation described in the
CC   disclosure of the invention
XX
SQ   Sequence 2308 BP; 268 A; 317 C; 813 G; 910 T; 0 U; 0 Other;

  Query Match            50.0%;  Score 30;  DB 6;  Length 2308;
  Best Local Similarity   72.2%;  Pred. No. 2;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||  ||||||||||  || ||      ||||  ||||||||||| || ||
Db       1222 GAAACCGAAACCGAAACGAAAACCGAAACCGAAACCGAAACCGAAACTAAAACC 1169


RESULT 13
ABQ39167
ID   ABQ39167 standard; DNA; 2308 BP.
XX
AC   ABQ39167;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 25758.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP010074.
XX
PR   01-SEP-2000; 2000DE-01043826.
PR   05-SEP-2000; 2000DE-01044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful for
PT   diagnosis and prognosis, comprises selective hybridization of amplicons
PT   from chemically treated DNA.
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one member,
CC   of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC   degree of hybridisation to both classes is determined from the label on
CC   the amplicon. From the ratio of labels hybridised to the two classes of
CC   oligomers, the degree of methylation is calculated. The method is used:
CC   (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC   and of a wide range of diseases, e.g. cancer, disorders of the central
CC   nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC   particularly by detecting mutations or single nucleotide polymorphisms
CC   (SNP's); and (ii) for differentiation of cell or tissue types and for
CC   investigating cell differentiation. The method allows the methylation
CC   status of many C residues to be determined simultaneously. ABQ13410-
CC   ABQ54121 represent genomic DNA sequences used to illustrate the method
CC   for determining the degree of cytosine methylation described in the
CC   disclosure of the invention
XX
SQ   Sequence 2308 BP; 910 A; 813 C; 317 G; 268 T; 0 U; 0 Other;

  Query Match            50.0%;  Score 30;  DB 6;  Length 2308;
  Best Local Similarity   72.2%;  Pred. No. 2;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||  ||||||||||  || ||      ||||  ||||||||||| || ||
Db       1087 GAAACCGAAACCGAAACGAAAACCGAAACCGAAACCGAAACCGAAACTAAAACC 1140


RESULT 14
ADO36037/c
ID   ADO36037 standard; DNA; 2775 BP.
XX
AC   ADO36037;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Novel mouse gene sequence #710.
XX
KW   mouse; murine; cancer; psoriasis; ulcerative colitis; inflammation;
KW   ischaemic heart disease; thrombosis; immune disorder; bacterial disorder;
KW   viral disorder; ds; gene.
XX
OS   Mus sp.
XX
PN   WO2004046310-A2.
XX
PD   03-JUN-2004.
XX
PF   24-OCT-2003; 2003WO-US033948.
XX
PR   15-NOV-2002; 2002US-0426916P.
PR   04-DEC-2002; 2002US-0431158P.
PR   05-DEC-2002; 2002US-0431445P.
PR   05-DEC-2002; 2002US-0431606P.
PR   09-JUN-2003; 2003US-0476621P.
PR   09-JUN-2003; 2003US-0476632P.
PR   08-JUL-2003; 2003US-0485217P.
PR   08-JUL-2003; 2003US-0485359P.
PR   08-AUG-2003; 2003US-0493332P.
PR   08-AUG-2003; 2003US-0493356P.
XX
```

```
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT,  Chu K,  Lee E,  Hestir K,  Hayashizaki Y,  Kamiya M;
XX
DR   WPI; 2004-431966/40.
XX
PT   New mouse nucleic acid molecules and polypeptides, useful for treating
PT   cancer, psoriasis, ulcerative colitis, inflammation, ischemic heart
PT   disease or thrombosis.
XX
PS   Claim 1; SEQ ID NO 710; 263pp; English.
XX
CC   The invention comprises 744 novel mouse DNA sequences (genes). The DNA
CC   sequences of the invention are useful for treating cancer, psoriasis,
CC   ulcerative colitis, inflammation, ischaemic heart disease, thrombosis,
CC   immune disorders, bacterial disorders and viral disorders. The present
CC   nucleic acid represents a mouse DNA sequence of the invention. NOTE: The
CC   present DNA sequence is not shown in the specification, but has been
CC   retrieved from the WIPO website.
XX
SQ   Sequence 2775 BP; 668 A; 617 C; 502 G; 988 T; 0 U; 0 Other;

  Query Match             50.0%;  Score 30;  DB 12;  Length 2775;
  Best Local Similarity   72.2%;  Pred. No. 2.1;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              ||||||||||||  ||||  |||       ||||||||||||  ||||  ||||
Db        114 GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAG 61


RESULT 15
ABL33564/c
ID   ABL33564 standard; DNA; 6061 BP.
XX
AC   ABL33564;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1537.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1537; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6061 BP; 1382 A; 147 C; 1835 G; 2697 T; 0 U; 0 Other;

  Query Match             50.0%;  Score 30;  DB 6;  Length 6061;
  Best Local Similarity   72.2%;  Pred. No. 2.4;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||  ||||||||||| ||| |||       ||  |||||||||||| || ||
Db       5091 GAAACCGAAACCGAAACCGAAACCCGAACCCGAACCGAAACCGAAACTAAAACC 5038


Search completed: March 30, 2006, 17:00:52
Job time : 1172 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2006, 17:12:58 ; Search time 2133 Seconds
(without alignments)
1316.092 Million cell updates/sec

Title: US-10-677-777-5
Perfect score: 60
Sequence: 1 ctcgggaaagggaaaccgaa.....ggaaaccgaaactgaagccc 60

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	32.8	54.7	418	8	DN645838	DN645838 G6646.17
c	2	31.6	52.7	366	5	BW458570	BW458570 BW458570
	3	31.6	52.7	508	5	BW434502	BW434502 BW434502
c	4	31	51.7	256	7	CO324686	CO324686 EK194524.
c	5	31	51.7	480	7	CO188844	CO188844 EK042163.
	6	30.6	51.0	142	9	AZ742375	AZ742375 RPCI-24-7
	7	30.6	51.0	975	9	CC297949	CC297949 CH261-178
	8	30	50.0	155	10	CE491929	CE491929 tigr-gss-
	9	30	50.0	294	9	AZ966113	AZ966113 2M0236A01
	10	30	50.0	417	9	AZ552622	AZ552622 RPCI-23-2
	11	30	50.0	428	5	BY261767	BY261767 BY261767
	12	30	50.0	468	9	AZ896327	AZ896327 RPCI-24-1
	13	30	50.0	547	9	BH119537	BH119537 RPCI-24-2
	14	30	50.0	551	7	CN689741	CN689741 E0282D11-
	15	30	50.0	583	9	AZ420391	AZ420391 1M0198P07
	16	30	50.0	593	10	CE370803	CE370803 tigr-gss-
	17	30	50.0	613	10	CE529392	CE529392 tigr-gss-
	18	30	50.0	700	8	DR431455	DR431455 nax43g06.
c	19	30	50.0	2775	4	AK051559	AK051559 Mus muscu
	20	29.6	49.3	768	8	CX636431	CX636431 UCRPT02_1
	21	29.6	49.3	778	9	BZ743227	BZ743227 OGEBW87TM
c	22	29.6	49.3	792	9	BZ743222	BZ743222 OGEBW87TC
c	23	29.6	49.3	819	8	CX636430	CX636430 UCRPT02_1
	24	29.4	49.0	585	10	CE283175	CE283175 tigr-gss-
c	25	29.4	49.0	603	10	AG913227	AG913227 Drosophil
	26	29.4	49.0	921	2	BG788523	BG788523 SEAUMC008
	27	29.4	49.0	1809	9	BZ580041	BZ580041 msh2_908.
c	28	29	48.3	117	11	CR057585	CR057585 Forward s
c	29	29	48.3	122	9	AZ809928	AZ809928 2M0074J08
	30	29	48.3	157	3	BM937104	BM937104 UI-M-CE0-
	31	29	48.3	186	10	CE788683	CE788683 tigr-gss-
c	32	29	48.3	226	7	CR527910	CR527910 CR527910
c	33	29	48.3	307	9	AQ982964	AQ982964 RPCI-23-3
	34	29	48.3	311	9	CE027875	CE027875 tigr-gss-
	35	29	48.3	353	3	BM496639	BM496639 IpCGBr2_4
c	36	29	48.3	354	5	BQ556001	BQ556001 H4037E04-
	37	29	48.3	378	9	CE030606	CE030606 tigr-gss-
	38	29	48.3	386	11	CR079775	CR079775 Forward s
c	39	29	48.3	419	10	CE844523	CE844523 tigr-gss-
	40	29	48.3	429	11	CR221714	CR221714 Reverse s
c	41	29	48.3	434	2	BI067609	BI067609 pgf1n.pk0
c	42	29	48.3	440	1	AI647275	AI647275 mq59a03.y
c	43	29	48.3	500	9	CE133347	CE133347 tigr-gss-
	44	29	48.3	540	7	CN668922	CN668922 A0873E07-
	45	29	48.3	543	3	BI566579	BI566579 RH36262.5

ALIGNMENTS

RESULT 1
DN645838
LOCUS DN645838 418 bp mRNA linear EST 28-MAR-2005
DEFINITION G6646.17 Exelixis Tribolium castaneum cDNA Library Tribolium castaneum cDNA clone G6646.17 5', mRNA sequence.
ACCESSION DN645838
VERSION DN645838.1 GI:61951239
KEYWORDS EST.
SOURCE Tribolium castaneum (red flour beetle)
ORGANISM Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tribolium.
REFERENCE 1 (bases 1 to 418)
AUTHORS Feder,J., Platt,D., Shao,A. and Swimmer,C.
TITLE A Tribolium castaneum mixed-stage, whole organism cDNA library
JOURNAL Unpublished (2005)
COMMENT Contact: Brown SJ
Divison of Biology, SJ Brown
Kansas State University
Ackert Hall, KSU, Manhattan, KS 66506, USA
Tel: 785 532 3935
Fax: 785 532 6653
Email: sjbrown@ksu.edu
POLYA=No.
FEATURES Location/Qualifiers
source 1. .418
/organism="Tribolium castaneum"
/mol_type="mRNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone="G6646.17"
/tissue_type="mixed"
/dev_stage="mixed larvae or adults"
/clone_lib="Exelixis Tribolium castaneum cDNA Library"
/note="Vector: SPORT 1"
ORIGIN

Query Match 54.7%; Score 32.8; DB 8; Length 418;
Best Local Similarity 76.9%; Pred. No. 3.8;
Matches 40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              |||| |||| ||| |||||||| |||||| ||||| ||| || ||  |||
Db         56 GGAAGGGGACACCAAAACTGAAACCCCTCCGGAAACAAAAAACGCAATCGAA 107
```

```
RESULT 2
BW458570/c
LOCUS       BW458570                366 bp    mRNA    linear   EST 10-JUN-2004
DEFINITION  BW458570 Nori Satoh unpublished cDNA library, juvenile whole animal
            Ciona intestinalis cDNA clone cijv029a13 3', mRNA sequence.
ACCESSION   BW458570
VERSION     BW458570.1  GI:48593358
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Satou,Y., Nakayama,A., Shin-i,T., Kohara,Y. and Satoh,N.
  TITLE     Expressed genes in Ciona intestinalis (2004b)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cijv029a13"
                     /tissue_type="whole animal"
                     /dev_stage="juvenile"
                     /clone_lib="Nori Satoh unpublished cDNA library, juvenile
                     whole animal"
ORIGIN

  Query Match             52.7%;  Score 31.6;  DB 5;  Length 366;
  Best Local Similarity   74.1%;  Pred. No. 9;
  Matches   40;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||| |||||| ||||| ||| |      |||||| |||||| ||||| ||| |
Db        336 GGAAACGGAAACGGAAACGGAAACGGAAACGGAAACGGAAACGGAAACGGAAAC 283


RESULT 3
BW434502
LOCUS       BW434502                508 bp    mRNA    linear   EST 03-JUN-2004
DEFINITION  BW434502 Nori Satoh unpublished cDNA library, juvenile whole animal
            Ciona intestinalis cDNA clone cijv029a13 5', mRNA sequence.
ACCESSION   BW434502
VERSION     BW434502.1  GI:48132510
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Satou,Y., Nakayama,A., Shin-i,T., Kohara,Y. and Satoh,N.
  TITLE     Expressed genes in Ciona intestinalis (2004b)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cijv029a13"
                     /tissue_type="whole animal"
                     /dev_stage="juvenile"
                     /clone_lib="Nori Satoh unpublished cDNA library, juvenile
                     whole animal"
ORIGIN

  Query Match             52.7%;  Score 31.6;  DB 5;  Length 508;
  Best Local Similarity   74.1%;  Pred. No. 9.3;
  Matches   40;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||| |||||| ||||| ||| |      |||||| |||||| ||||| ||| |
Db        187 GGAAACGGAAACGGAAACGGAAACGGAAACGGAAACGGAAACGGAAACGGAAAC 240


RESULT 4
CO324686/c
LOCUS       CO324686                256 bp    mRNA    linear   EST 28-JUN-2004
DEFINITION  EK194524.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK194524 5, mRNA sequence.
ACCESSION   CO324686
VERSION     CO324686.1  GI:49383120
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: EK.1945  row: B  column: 12
            High quality sequence stop: 195.
FEATURES             Location/Qualifiers
     source          1. .256
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK194524"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

  Query Match             51.7%;  Score 31;  DB 7;  Length 256;
  Best Local Similarity   72.7%;  Pred. No. 14;
  Matches   40;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      ||||| | ||||  |||| |||| |
Db        190 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 136


RESULT 5
CO188844/c
LOCUS       CO188844                480 bp    mRNA    linear   EST 21-JUN-2004
DEFINITION  EK042163.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK042163 5, mRNA sequence.
ACCESSION   CO188844
VERSION     CO188844.1  GI:49000019
KEYWORDS    EST.
```

```
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon one or more reads of this clone where vector sequence
            was present at both ends, this clone has been determined to contain
            contain a cDNA insert on the order of 600-1000 bases.
            Plate: EK.421  row: F  column: 3
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK042163"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

  Query Match            51.7%;  Score 31;  DB 7;  Length 480;
  Best Local Similarity  72.7%;  Pred. No. 14;
  Matches   40;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      ||||| | ||||  |||| |||| |
Db        190 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 136


RESULT 6
AZ742375
LOCUS       AZ742375                142 bp    DNA     linear   GSS 25-JAN-2001
DEFINITION  RPCI-24-74F5.TJ RPCI-24 Mus musculus genomic clone RPCI-24-74F5,
            genomic survey sequence.
ACCESSION   AZ742375
VERSION     AZ742375.1  GI:12521185
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 142)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-74F5.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 74  row: F  column: 5
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .142
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-74F5"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
ORIGIN

  Query Match            51.0%;  Score 30.6;  DB 9;  Length 142;
  Best Local Similarity  73.6%;  Pred. No. 17;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              |||||| |||||  ||||  |||  |    ||||||||||||  ||||| |||
Db          8 GGGAAACGGAAAGGGAAAGGGAAAGCGAAAGGGAAAGGGAAAGGGAAACGGAA 60


RESULT 7
CC297949
LOCUS       CC297949                975 bp    DNA     linear   GSS 13-MAY-2003
DEFINITION  CH261-178B1_Sp6.1 CH261 Gallus gallus genomic clone CH261-178B1,
            genomic survey sequence.
ACCESSION   CC297949
VERSION     CC297949.1  GI:30669390
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 975)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 25
            High quality sequence stop: 756.
FEATURES             Location/Qualifiers
     source          1. .975
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-178B1"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN
```

```
  Query Match             51.0%;  Score 30.6;  DB 9;  Length 975;
  Best Local Similarity   73.6%;  Pred. No. 21;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              |||||||||||| ||||  |||     | ||||||||||| ||||  |||
Db        420 GGGAAAGGGAAAGGGAAAGGGAAATTTTTTGGGAAAGGGAAAGGGAAAGGGAA 472


RESULT 8
CE491929
LOCUS       CE491929                155 bp    DNA     linear   GSS 28-SEP-2003
DEFINITION  tigr-gss-dog-17000309547760 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE491929
VERSION     CE491929.1  GI:36808710
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 155)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .155
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             50.0%;  Score 30;  DB 10;  Length 155;
  Best Local Similarity   72.2%;  Pred. No. 27;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              |||||||||||| ||||  |||       |||||||||||| ||||  ||||
Db         85 GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAG 138


RESULT 9
AZ966113
LOCUS       AZ966113                294 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0236A01R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0236A01 R, genomic survey sequence.
ACCESSION   AZ966113
VERSION     AZ966113.1  GI:13837340
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0236  row: A  column: 01
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 294.
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0236A01"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

  Query Match             50.0%;  Score 30;  DB 9;  Length 294;
  Best Local Similarity   72.2%;  Pred. No. 29;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              |||||||||||| ||||  |||       |||||||||||| ||||  ||||
Db         24 GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAG 77


RESULT 10
AZ552622
LOCUS       AZ552622                417 bp    DNA     linear   GSS 20-NOV-2000
DEFINITION  RPCI-23-210N2.TV RPCI-23 Mus musculus genomic clone RPCI-23-210N2,
            genomic survey sequence.
ACCESSION   AZ552622
VERSION     AZ552622.1  GI:11232122
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
```

```
            Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 210  row: N  column: 2
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-210N2"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
ORIGIN

  Query Match           50.0%;  Score 30;  DB 9;  Length 417;
  Best Local Similarity   72.2%;  Pred. No. 30;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              ||||||||||||  ||||  |||       ||||||||||||  ||||  ||||
Db        108 GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAG 161


RESULT 11
BY261767
LOCUS       BY261767                428 bp    mRNA    linear   EST 10-DEC-2002
DEFINITION  BY261767 RIKEN full-length enriched, visual cortex Mus musculus
            cDNA clone K330308L17 5', mRNA sequence.
ACCESSION   BY261767
VERSION     BY261767.1  GI:26443279
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  PUBMED    12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Michela Fagiolini and Takao K. Hensch (
            Laboratory for Neuronal Circuit Development Brain Science Institute
            RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
            assistance we gratefully acknowledge. Please visit our web site
            (http://genome.gsc.riken.go.jp) for further details.
FEATURES             Location/Qualifiers
     source          1. .428
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K330308L17"
                     /tissue_type="visual cortex"
                     /clone_lib="RIKEN full-length enriched, visual cortex"
ORIGIN

  Query Match           50.0%;  Score 30;  DB 5;  Length 428;
  Best Local Similarity   72.2%;  Pred. No. 30;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | | ||  |||| | ||  ||  ||  |||||||  |||||||| |    ||||  ||  ||
Db        119 GTGTAATGGAACCAGACACGGATGCCCCTCAGGAAAGGGGAGATGAAAATGCAG 172
```

```
RESULT 12
AZ896327
LOCUS       AZ896327                468 bp    DNA     linear   GSS 05-MAR-2001
DEFINITION  RPCI-24-171J3.TV RPCI-24 Mus musculus genomic clone RPCI-24-171J3,
            genomic survey sequence.
ACCESSION   AZ896327
VERSION     AZ896327.1  GI:13215272
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-171J3.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 171  row: J  column: 3
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-171J3"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
ORIGIN

  Query Match             50.0%;  Score 30;  DB 9;  Length 468;
  Best Local Similarity   72.2%;  Pred. No. 30;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              ||||||||||||  ||||  |||        ||||||||||||  ||||  ||||
Db        375 GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAG 428


RESULT 13
BH119537
LOCUS       BH119537                547 bp    DNA     linear   GSS 19-JUL-2001
DEFINITION  RPCI-24-299K7.TJ RPCI-24 Mus musculus genomic clone RPCI-24-299K7,
            genomic survey sequence.
ACCESSION   BH119537
VERSION     BH119537.1  GI:14962981
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-299K7.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 299  row: K  column: 7
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-299K7"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
ORIGIN

  Query Match             50.0%;  Score 30;  DB 9;  Length 547;
  Best Local Similarity   72.2%;  Pred. No. 31;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              ||||||||||||  ||||  |||        ||||||||||||  ||||  ||||
Db        354 GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAG 407


RESULT 14
CN689741
LOCUS       CN689741                551 bp    mRNA    linear   EST 17-MAY-2004
DEFINITION  E0282D11-5 NIA Mouse Embryonic Stem (ES) cell (Lif-, 48 h, high
            density) cDNA library (Long) Mus musculus cDNA clone NIA:E0282D11
            IMAGE:30857230 5', mRNA sequence.
ACCESSION   CN689741
VERSION     CN689741.1  GI:47456187
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Sharov,A.A., Piao,Y., Matoba,R., Dudekula,D.B., Qian,Y.,
            VanBuren,V., Falco,G., Martin,P.R., Stagg,C.A., Bassey,U.C.,
            Wang,Y., Carter,M.G., Hamatani,T., Aiba,K., Akutsu,H., Sharova,L.,
            Tanaka,T.S., Kimber,W.L., Yoshikawa,T., Jaradat,S.A., Pantano,S.,
            Nagaraja,R., Boheler,K.R., Taub,D., Hodes,R.J., Longo,D.L.,
            Schlessinger,D., Keller,J., Klotz,E., Kelsoe,G., Umezawa,A.,
            Vescovi,A.L., Rossant,J., Kunath,T., Hogan,B.L., Curci,A.,
```

```
            D'Urso,M., Kelso,J., Hide,W. and Ko,M.S.
  TITLE     Transcriptome analysis of mouse stem cells and early embryos
  JOURNAL   PLoS Biol. 1 (3), 410-419 (2003)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: E0282  row: D  column: 11
            Seq primer: M13 Reverse
            High quality sequence stop: 551
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129Sv/EvTac"
                     /db_xref="niaEST:E0282D11-5"
                     /db_xref="taxon:10090"
                     /clone="NIA:E0282D11 IMAGE:30857230"
                     /sex="Male"
                     /tissue_type="Embryonic Stem Cell"
                     /cell_line="129.3 ES cells"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse Embryonic Stem (ES) cell (Lif-, 48
                     h, high density) cDNA library (Long)"
                     /note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
                     Site_2: NotI; Mouse cDNA project by the Laboratory of
                     Genetics, National Institute on Aging (NIA), Intramural
                     Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
                     This is a long-transcript enriched cDNA library (Ref.
                     Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). ES
                     cells were plated at density 3x104/cm2, on gelatin-coated
                     plates and cultured for 48 hrs at 37 OC, 5% CO2. Culture
                     medium: DMEM supplemented with 15% FBS, 2 mM L-glutamine,
                     0.1 mM NEAA, 1mM Sodium pyruvate, 0.1 mM
                     beta-mercaptoethanol, 100 U/ml penicillin, and 100 ug/ml
                     streptomycin. Double-stranded cDNAs were synthesized with
                     an Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     2ug of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
                     vector. The DH10B E. coli host was transformed with the
                     ligation mixture by the standard chemical method. The
                     average insert size is about 3.4 kb. The library was
                     constructed by Yulan Piao."
ORIGIN

  Query Match             50.0%;  Score 30;  DB 7;  Length 551;
  Best Local Similarity   72.2%;  Pred. No. 31;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | | ||  ||| | || ||||| ||||||| ||||| || ||  |||| || ||
Db        196 GTGTAATTGAACCAGACACTGATGCCCCTCAGGAAATGGGAAATGAAAATGCAG 249


RESULT 15
AZ420391
LOCUS       AZ420391                583 bp    DNA     linear   GSS 03-OCT-2000
DEFINITION  1M0198P07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0198P07 F, genomic survey sequence.
ACCESSION   AZ420391
VERSION     AZ420391.1  GI:10544404
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0198  row: P  column: 07
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 583.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0198P07"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

  Query Match             50.0%;  Score 30;  DB 9;  Length 583;
  Best Local Similarity   72.2%;  Pred. No. 31;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              ||||||||||||  ||||  |||       ||||||||||||  ||||  ||||
Db        447 GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAG 500


Search completed: March 30, 2006, 18:17:08
Job time : 2137 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2006, 17:41:42 ; Search time 96 Seconds
(without alignments)
1110.976 Million cell updates/sec

Title: US-10-677-777-5
Perfect score: 60
Sequence: 1 ctcgggaaagggaaaccgaa.....ggaaaccgaaactgaagccc 60

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	31	51.7	466	3	US-09-270-767-8787	Sequence 8787, Ap
c	2	31	51.7	466	3	US-09-270-767-24069	Sequence 24069, A
	3	30	50.0	33	2	US-08-078-222B-8	Sequence 8, Appli
	4	30	50.0	33	2	US-08-661-330A-8	Sequence 8, Appli
	5	30	50.0	33	3	US-09-038-217A-8	Sequence 8, Appli
	6	30	50.0	33	3	US-09-447-034-8	Sequence 8, Appli
	7	30	50.0	34	3	US-09-079-030-229	Sequence 229, App
	8	30	50.0	40	2	US-08-141-499A-14	Sequence 14, Appl
	9	30	50.0	40	2	US-08-467-940-14	Sequence 14, Appl
	10	30	50.0	40	2	US-08-633-772-14	Sequence 14, Appl
c	11	26.8	44.7	23193	3	US-09-949-016-17215	Sequence 17215, A
	12	26.4	44.0	1296	3	US-09-248-796A-11008	Sequence 11008, A
c	13	26.2	43.7	527	3	US-09-270-767-8495	Sequence 8495, Ap
c	14	26.2	43.7	527	3	US-09-270-767-23777	Sequence 23777, A
	15	26	43.3	561	3	US-09-248-796A-5652	Sequence 5652, Ap
	16	26	43.3	3337	2	US-08-072-610-1	Sequence 1, Appli
	17	26	43.3	3337	2	US-08-719-822B-1	Sequence 1, Appli
	18	26	43.3	3337	3	US-09-092-458-1	Sequence 1, Appli
	19	26	43.3	3337	3	US-08-719-821C-1	Sequence 1, Appli
	20	25.8	43.0	1809	3	US-09-248-796A-5736	Sequence 5736, Ap
	21	25.2	42.0	2609	3	US-10-104-047-1365	Sequence 1365, Ap
	22	25.2	42.0	6158	3	US-09-799-451-897	Sequence 897, App
	23	25.2	42.0	6429	3	US-10-144-198-3	Sequence 3, Appli
	24	25.2	42.0	6726	3	US-10-144-198-1	Sequence 1, Appli
	25	24.8	41.3	282	3	US-09-270-767-4609	Sequence 4609, Ap
	26	24.8	41.3	282	3	US-09-270-767-19891	Sequence 19891, A
c	27	24.6	41.0	195	3	US-09-513-999C-18180	Sequence 18180, A
c	28	24.4	40.7	84761	3	US-09-949-016-11919	Sequence 11919, A
c	29	24.4	40.7	84763	3	US-09-949-016-13914	Sequence 13914, A
c	30	24.2	40.3	273	3	US-09-270-767-9738	Sequence 9738, Ap
c	31	24.2	40.3	273	3	US-09-270-767-25020	Sequence 25020, A
	32	24.2	40.3	1274	3	US-09-270-767-10133	Sequence 10133, A
	33	24.2	40.3	36203	3	US-09-949-016-15430	Sequence 15430, A
c	34	24.2	40.3	44715	3	US-09-949-016-16353	Sequence 16353, A
c	35	24	40.0	28	3	US-08-956-653A-26	Sequence 26, Appl
	36	24	40.0	387	3	US-09-902-540-2794	Sequence 2794, Ap
	37	24	40.0	507	3	US-08-991-789A-253	Sequence 253, App
	38	24	40.0	507	3	US-09-062-451-253	Sequence 253, App
	39	24	40.0	507	3	US-09-289-198-253	Sequence 253, App
	40	24	40.0	507	3	US-09-429-755-253	Sequence 253, App
	41	24	40.0	507	3	US-09-699-295-253	Sequence 253, App
	42	24	40.0	507	3	US-09-534-825A-253	Sequence 253, App
c	43	24	40.0	601	3	US-09-949-016-80571	Sequence 80571, A
c	44	24	40.0	2350	3	US-10-104-047-1721	Sequence 1721, Ap
	45	24	40.0	15644	3	US-09-902-540-1133	Sequence 1133, Ap

ALIGNMENTS

RESULT 1
US-09-270-767-8787/c
; Sequence 8787, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8787
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-8787

Query Match 51.7%; Score 31; DB 3; Length 466;
Best Local Similarity 72.7%; Pred. No. 0.22;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      ||||| | ||||  |||| |||| |
Db        189 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 135
```

RESULT 2
US-09-270-767-24069/c
; Sequence 24069, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24069
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24069

Query Match 51.7%; Score 31; DB 3; Length 466;
Best Local Similarity 72.7%; Pred. No. 0.22;

```
  Matches   40;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||| ||||| |||||| |||||      |||||  | ||||  |||| |||| |
Db        189 GGAAATGGAAATCGAAACAGAAGCGGAAACGGAAACGAAAACGAAAACGGAAGAC 135


RESULT 3
US-08-078-222B-8
; Sequence 8, Application US/08078222B
; Patent No. 5604115
;  GENERAL INFORMATION:
;    APPLICANT:  The Rockefeller University
;    TITLE OF INVENTION:  Liver Enriched Transcription Factor
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  James E. Darnell, Jr.
;      STREET:  96 Edgewood Avenue
;      CITY:  Larchmont
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10538
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.45 Mb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/078,222B
;      FILING DATE:  October 28, 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson, David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-030CIP PCT US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 487-5800
;      TELEFAX:  (201) 343-1684
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    ANTI-SENSE:  NO
US-08-078-222B-8

  Query Match             50.0%;  Score 30;  DB 2;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.28;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 4
US-08-661-330A-8
; Sequence 8, Application US/08661330A
; Patent No. 5849485
;  GENERAL INFORMATION:
;    APPLICANT:  Sladek, Frances M.
;    APPLICANT:  Zhong, Weimin
;    APPLICANT:  Darnell, Jr., James F.
;    TITLE OF INVENTION:  LIVER ENRICHED TRANSCRIPTION FACTOR
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/661,330A
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-030A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
US-08-661-330A-8

  Query Match             50.0%;  Score 30;  DB 2;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.28;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 5
US-09-038-217A-8
; Sequence 8, Application US/09038217A
; Patent No. 6025196
;  GENERAL INFORMATION:
;    APPLICANT:  Sladek, Frances M.
;    APPLICANT:  Zhong, Weimin
;    APPLICANT:  Darnell, Jr., James F.
;    TITLE OF INVENTION:  LIVER ENRICHED TRANSCRIPTION FACTOR
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/038,217A
;      FILING DATE:  March 11, 1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-030B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  33 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
US-09-038-217A-8

  Query Match             50.0%;  Score 30;  DB 3;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.28;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 6
US-09-447-034-8
; Sequence 8, Application US/09447034
; Patent No. 6500672
;   GENERAL INFORMATION:
;        APPLICANT: Sladek, Frances M.
;                   Zhong, Weimin
;                   Darnell, Jr., James F.
;        TITLE OF INVENTION: LIVER ENRICHED TRANSCRIPTION FACTOR
;        NUMBER OF SEQUENCES: 18
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: David A. Jackson, Esq.
;             STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                     Floor
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/447,034
;             FILING DATE: 22-No. 6500672-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/038,217
;             FILING DATE: March 11, 1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 600-1-030B
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201-487-5800
;             TELEFAX: 201-343-1684
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 33 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        ANTI-SENSE: NO
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-447-034-8

  Query Match             50.0%;  Score 30;  DB 3;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 0.28;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 7
US-09-079-030-229
; Sequence 229, Application US/09079030
; Patent No. 6635623
;  GENERAL INFORMATION:
;    APPLICANT:  Guevera, Jr., Juan G.
;    APPLICANT:  Hoogeveen, Ron C.
;    APPLICANT:  Moore, Paul J.
;    TITLE OF INVENTION:  LIPOPROTEINS AS NUCLEIC ACID DELIVERY
;    TITLE OF INVENTION:  VECTORS FOR TRANSFECTION OF EUKARYOTIC CELLS
;    NUMBER OF SEQUENCES:  229
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/079,030
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMillian, Nabeela R.
;      REGISTRATION NUMBER:  P-43,363
;      REFERENCE/DOCKET NUMBER:  ARAG:003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  229:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  34 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-079-030-229

  Query Match             50.0%;  Score 30;  DB 3;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 0.28;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 8
US-08-141-499A-14
; Sequence 14, Application US/08141499A
; Patent No. 5463023
;  GENERAL INFORMATION:
;    APPLICANT:  Schreiber, Robert D.
;    APPLICANT:  Farrar, Michael A.
;    APPLICANT:  Greenlund, Andrew C.
;    TITLE OF INVENTION:  Composition for Inhibition of
;    TITLE OF INVENTION:  Intracellular Transcription Factor and Method Therefor
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rogers, Howell & Haferkamp
;      STREET:  7733 Forsyth Blvd Suite 1400
;      CITY:  St. Louis
;      STATE:  MO
;      COUNTRY:  USA
;      ZIP:  63105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/141,499A
;       FILING DATE:  22-OCT-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Holland, Donald R.
;       REGISTRATION NUMBER:  35,197
;       REFERENCE/DOCKET NUMBER:  939102
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  314-727-5188
;       TELEFAX:  314-727-6092
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  40 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-141-499A-14

  Query Match             50.0%;  Score 30;  DB 2;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 0.29;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          9 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 38


RESULT 9
US-08-467-940-14
; Sequence 14, Application US/08467940
; Patent No. 5582999
;  GENERAL INFORMATION:
;    APPLICANT:  Schreiber, Robert D.
;    APPLICANT:  Farrar, Michael A.
;    APPLICANT:  Greenlund, Andrew C.
;    TITLE OF INVENTION:  Composition for Inhibition of
;    TITLE OF INVENTION:  Intracellular Transcription Factor and Method Therefor
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rogers, Howell & Haferkamp
;      STREET:  7733 Forsyth Blvd Suite 1400
;      CITY:  St. Louis
;      STATE:  MO
;      COUNTRY:  USA
;      ZIP:  63105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,940
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Holland, Donald R.
;      REGISTRATION NUMBER:  35,197
;      REFERENCE/DOCKET NUMBER:  952999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  314-727-5188
;      TELEFAX:  314-727-6092
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-467-940-14

  Query Match             50.0%;  Score 30;  DB 2;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 0.29;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          9 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 38


RESULT 10
US-08-633-772-14
; Sequence 14, Application US/08633772
; Patent No. 5731155
;  GENERAL INFORMATION:
;    APPLICANT:  Schreiber, Robert D.
;    APPLICANT:  Farrar, Michael A.
;    APPLICANT:  Greenlund, Andrew C.
;    TITLE OF INVENTION:  Compositions for Inhibition of
;    TITLE OF INVENTION:  Intracellular Transcription Factors and Methods
;    TITLE OF INVENTION:  Therefor
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howell & Haferkamp, LC
;      STREET:  7733 Forsyth Boulevard
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/633,772
;      FILING DATE:  21-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/141,499
;      FILING DATE:  22-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Holland, Donald R.
;      REGISTRATION NUMBER:  35197
;      REFERENCE/DOCKET NUMBER:  964635
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  314-727-5188
;      TELEFAX:  314-727-6092
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-633-772-14

  Query Match             50.0%;  Score 30;  DB 2;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 0.29;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          9 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 38


RESULT 11
US-09-949-016-17215/c
```

; Sequence 17215, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17215
; LENGTH: 23193
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(23193)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17215

Query Match 44.7%; Score 26.8; DB 3; Length 23193;
Best Local Similarity 73.9%; Pred. No. 13;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAA 50
Db 8378 GGAAAGGGAAAGGGAAAGGGAAAGGGCAAGGGAAAGGGAAAGGGAA 8333

RESULT 12
US-09-248-796A-11008
; Sequence 11008, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 11008
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-11008

Query Match 44.0%; Score 26.4; DB 3; Length 1296;
Best Local Similarity 69.2%; Pred. No. 9.5;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
Db 1183 GAAACGGAAACGGAAACGGAAACCGAGAATGAAACAGGAACAGGAACAGAAG 1234

RESULT 13
US-09-270-767-8495/c
; Sequence 8495, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8495
; LENGTH: 527
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-8495

Query Match 43.7%; Score 26.2; DB 3; Length 527;
Best Local Similarity 67.3%; Pred. No. 9.2;
Matches 37; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
Db 252 GGAAACGGAAATGGGAACAGAAACAAAAACAGAAATGGAAACCGAAATCGTAACC 198

RESULT 14
US-09-270-767-23777/c
; Sequence 23777, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23777
; LENGTH: 527
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-23777

Query Match 43.7%; Score 26.2; DB 3; Length 527;
Best Local Similarity 67.3%; Pred. No. 9.2;
Matches 37; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 5 GGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
Db 252 GGAAACGGAAATGGGAACAGAAACAAAAACAGAAATGGAAACCGAAATCGTAACC 198

RESULT 15
US-09-248-796A-5652
; Sequence 5652, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5652
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5652

Query Match 43.3%; Score 26; DB 3; Length 561;
Best Local Similarity 65.5%; Pred. No. 11;
Matches 38; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||   |||   ||||| |||||||| || ||    |||   ||| | |||  ||||||
Db        494 CTGCTGAAGATGAAACTGAAACTGATGCTCCAGTTGAAGCTGAAGCTGAAGTTGAAGC 551
```

Search completed: March 30, 2006, 18:18:47
Job time : 97 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2006, 19:22:21 ; Search time 452 Seconds
(without alignments)
1097.705 Million cell updates/sec

Title: US-10-677-777-5
Perfect score: 60
Sequence: 1 ctcgggaaagggaaaccgaa.....ggaaaccgaaactgaagccc 60

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	60	100.0	60	8	US-10-677-777-5	Sequence 5, Appli
2	60	100.0	4412	8	US-10-677-777-3	Sequence 3, Appli
3	60	100.0	5849	8	US-10-677-777-1	Sequence 1, Appli
c 4	30.6	51.0	1413	8	US-10-363-345A-19419	Sequence 19419, A
5	30.6	51.0	1413	8	US-10-363-345A-19420	Sequence 19420, A
c 6	30.6	51.0	1413	9	US-10-363-483A-19419	Sequence 19419, A
7	30.6	51.0	1413	9	US-10-363-483A-19420	Sequence 19420, A
c 8	30.6	51.0	1418	8	US-10-363-345A-19433	Sequence 19433, A
9	30.6	51.0	1418	8	US-10-363-345A-19434	Sequence 19434, A
c 10	30.6	51.0	1418	9	US-10-363-483A-19433	Sequence 19433, A
11	30.6	51.0	1418	9	US-10-363-483A-19434	Sequence 19434, A
12	30	50.0	33	5	US-10-215-597-8	Sequence 8, Appli
13	30	50.0	34	9	US-10-656-053B-229	Sequence 229, App
c 14	30	50.0	2308	8	US-10-363-345A-25757	Sequence 25757, A
15	30	50.0	2308	8	US-10-363-345A-25758	Sequence 25758, A
c 16	30	50.0	2308	9	US-10-363-483A-25757	Sequence 25757, A
17	30	50.0	2308	9	US-10-363-483A-25758	Sequence 25758, A
c 18	30	50.0	6061	6	US-10-311-455-1537	Sequence 1537, Ap
c 19	30	50.0	6061	7	US-10-221-613-271	Sequence 271, App
c 20	30	50.0	6510	6	US-10-311-455-367	Sequence 367, App
c 21	30	50.0	6510	7	US-10-221-613-71	Sequence 71, Appl
c 22	30	50.0	6510	7	US-10-221-714A-369	Sequence 369, App
c 23	30	50.0	69515	5	US-10-087-192-1219	Sequence 1219, Ap
24	30	50.0	410846	9	US-10-481-613-1	Sequence 1, Appli
c 25	29	48.3	1067	8	US-10-363-345A-22667	Sequence 22667, A
26	29	48.3	1067	8	US-10-363-345A-22668	Sequence 22668, A
c 27	29	48.3	1067	9	US-10-363-483A-22667	Sequence 22667, A
28	29	48.3	1067	9	US-10-363-483A-22668	Sequence 22668, A
c 29	29	48.3	2365	8	US-10-363-345A-3659	Sequence 3659, Ap
30	29	48.3	2365	8	US-10-363-345A-3660	Sequence 3660, Ap
c 31	29	48.3	2365	9	US-10-363-483A-3659	Sequence 3659, Ap
32	29	48.3	2365	9	US-10-363-483A-3660	Sequence 3660, Ap
c 33	29	48.3	2395	8	US-10-363-345A-311	Sequence 311, App
34	29	48.3	2395	8	US-10-363-345A-312	Sequence 312, App
c 35	29	48.3	2395	9	US-10-363-483A-311	Sequence 311, App
36	29	48.3	2395	9	US-10-363-483A-312	Sequence 312, App
c 37	29	48.3	14693	8	US-10-719-993-6814	Sequence 6814, Ap
c 38	29	48.3	32343	8	US-10-719-993-6773	Sequence 6773, Ap
c 39	29	48.3	34460	5	US-10-087-192-115	Sequence 115, App
40	29	48.3	99924	6	US-10-085-117-115	Sequence 115, App
41	29	48.3	110459	10	US-11-013-314-47	Sequence 47, Appl
42	28.8	48.0	13497	7	US-10-712-124-67	Sequence 67, Appl
43	28.8	48.0	13497	10	US-11-097-143-2576	Sequence 2576, Ap
c 44	28.8	48.0	16489	10	US-11-097-143-2575	Sequence 2575, Ap
c 45	28.4	47.3	642	8	US-10-363-345A-40457	Sequence 40457, A

ALIGNMENTS

RESULT 1
US-10-677-777-5
; Sequence 5, Application US/10677777
; Publication No. US20040235157A1
; GENERAL INFORMATION:
; APPLICANT: TOVEY, Michael
; APPLICANT: LALLEMAND, Christophe
; TITLE OF INVENTION: GENE REPORTER ASSAY, KIT, AND CELLS FOR DETERMINING THE PRESENCE
; TITLE OF INVENTION: AND/OR THE LEVEL OF A MOLECULE THAT ACTIVATES SIGNAL TRANSDUCTI
; TITLE OF INVENTION: ACTIVITY OF A CELL SURFACE PROTEIN
; FILE REFERENCE: TOVEY=5A
; CURRENT APPLICATION NUMBER: US/10/677,777
; CURRENT FILING DATE: 2003-10-03
; PRIOR APPLICATION NUMBER: 60/415,818
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-677-777-5

Query Match 100.0%; Score 60; DB 8; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.5e-10;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60

RESULT 2
US-10-677-777-3
; Sequence 3, Application US/10677777
; Publication No. US20040235157A1
; GENERAL INFORMATION:
; APPLICANT: TOVEY, Michael
; APPLICANT: LALLEMAND, Christophe
; TITLE OF INVENTION: GENE REPORTER ASSAY, KIT, AND CELLS FOR DETERMINING THE PRESENCE
; TITLE OF INVENTION: AND/OR THE LEVEL OF A MOLECULE THAT ACTIVATES SIGNAL TRANSDUCTI
; TITLE OF INVENTION: ACTIVITY OF A CELL SURFACE PROTEIN
; FILE REFERENCE: TOVEY=5A
; CURRENT APPLICATION NUMBER: US/10/677,777

```
;  CURRENT FILING DATE:  2003-10-03
;  PRIOR APPLICATION NUMBER: 60/415,818
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
;   LENGTH: 4412
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Synthetic
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (358)..(1077)
US-10-677-777-3

  Query Match             100.0%;  Score 60;  DB 8;  Length 4412;
  Best Local Similarity   100.0%;  Pred. No. 4e-10;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         30 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 89


RESULT 3
US-10-677-777-1
; Sequence 1, Application US/10677777
; Publication No. US20040235157A1
; GENERAL INFORMATION:
;  APPLICANT: TOVEY, Michael
;  APPLICANT:  LALLEMAND, Christophe
;  TITLE OF INVENTION: GENE REPORTER ASSAY, KIT, AND CELLS FOR DETERMINING THE PRESENCE
;  TITLE OF INVENTION:  AND/OR THE LEVEL OF A MOLECULE THAT ACTIVATES SIGNAL TRANSDUCTIO
;  TITLE OF INVENTION:  ACTIVITY OF A CELL SURFACE PROTEIN
;  FILE REFERENCE: TOVEY=5A
;  CURRENT APPLICATION NUMBER: US/10/677,777
;  CURRENT FILING DATE:  2003-10-03
;  PRIOR APPLICATION NUMBER: 60/415,818
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 5849
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Synthetic
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (328)..(1980)
US-10-677-777-1

  Query Match             100.0%;  Score 60;  DB 8;  Length 5849;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-10;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         38 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCCC 97


RESULT 4
US-10-363-345A-19419/c
; Sequence 19419, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 19419
;   LENGTH: 1413
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 19419
US-10-363-345A-19419

  Query Match             51.0%;  Score 30.6;  DB 8;  Length 1413;
  Best Local Similarity   73.6%;  Pred. No. 1.6;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  |||||||||||| ||| ||      ||||  |||||||||||| ||| |
Db       1161 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 1109


RESULT 5
US-10-363-345A-19420
; Sequence 19420, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 19420
;   LENGTH: 1413
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 19420
US-10-363-345A-19420

  Query Match             51.0%;  Score 30.6;  DB 8;  Length 1413;
  Best Local Similarity   73.6%;  Pred. No. 1.6;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  |||||||||||| ||| ||      ||||  |||||||||||| ||| |
Db        253 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 305


RESULT 6
US-10-363-483A-19419/c
; Sequence 19419, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 19419
;   LENGTH: 1413
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
;   OTHER INFORMATION: CpG-island No: 19419
US-10-363-483A-19419

  Query Match             51.0%;  Score 30.6;  DB 9;  Length 1413;
  Best Local Similarity   73.6%;  Pred. No. 1.6;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  |||||||||||| ||| ||      ||||  |||||||||||| ||| |
Db       1161 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 1109


RESULT 7
US-10-363-483A-19420
; Sequence 19420, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 19420
;   LENGTH: 1413
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 19420
US-10-363-483A-19420

  Query Match             51.0%;  Score 30.6;  DB 9;  Length 1413;
  Best Local Similarity   73.6%;  Pred. No. 1.6;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  |||||||||||| ||| ||      ||||  |||||||||||| ||| |
Db        253 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 305


RESULT 8
US-10-363-345A-19433/c
; Sequence 19433, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 19433
;   LENGTH: 1418
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 19433
US-10-363-345A-19433

  Query Match             51.0%;  Score 30.6;  DB 8;  Length 1418;
  Best Local Similarity   73.6%;  Pred. No. 1.6;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  |||||||||||| ||| ||      ||||  |||||||||||| ||| |
Db       1075 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 1023


RESULT 9
US-10-363-345A-19434
; Sequence 19434, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 19434
;   LENGTH: 1418
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 19434
US-10-363-345A-19434

  Query Match             51.0%;  Score 30.6;  DB 8;  Length 1418;
  Best Local Similarity   73.6%;  Pred. No. 1.6;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  |||||||||||| ||| ||      ||||  |||||||||||| ||| |
Db        344 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 396


RESULT 10
US-10-363-483A-19433/c
; Sequence 19433, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 19433
;   LENGTH: 1418
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 19433
US-10-363-483A-19433

  Query Match             51.0%;  Score 30.6;  DB 9;  Length 1418;
  Best Local Similarity   73.6%;  Pred. No. 1.6;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  |||||||||||| ||| ||      ||||  |||||||||||| ||| |
Db       1075 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 1023


RESULT 11
US-10-363-483A-19434
; Sequence 19434, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
```

```
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 19434
;   LENGTH: 1418
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 19434
US-10-363-483A-19434

  Query Match             51.0%;  Score 30.6;  DB 9;  Length 1418;
  Best Local Similarity   73.6%;  Pred. No. 1.6;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              ||||  ||||||||||| ||| ||       ||||  |||||||||||| ||| |
Db        344 GAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAACCGAAAC 396


RESULT 12
US-10-215-597-8
; Sequence 8, Application US/10215597
; Publication No. US20030108912A1
;   GENERAL INFORMATION:
;        APPLICANT: Sladek, Frances M.
;                   Zhong, Weimin
;                   Darnell, Jr., James F.
;        TITLE OF INVENTION: LIVER ENRICHED TRANSCRIPTION FACTOR
;        NUMBER OF SEQUENCES: 18
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: David A. Jackson, Esq.
;             STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                     Floor
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/215,597
;             FILING DATE: 09-Aug-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/447,034
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 600-1-030B
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201-487-5800
;             TELEFAX: 201-343-1684
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 33 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        ANTI-SENSE: NO
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-215-597-8

  Query Match             50.0%;  Score 30;  DB 5;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 13
US-10-656-053B-229
; Sequence 229, Application US/10656053B
; Publication No. US20050215499A1
; GENERAL INFORMATION:
;  APPLICANT: GUEVERA, JR., JUAN G.
;  APPLICANT:  HOOGEVEEN, RON C.
;  APPLICANT:  MOORE, PAUL J.
;  TITLE OF INVENTION: LIPOPROTEINS AS NUCLEIC ACID VECTORS
;  FILE REFERENCE: ARAG:003USD1
;  CURRENT APPLICATION NUMBER: US/10/656,053B
;  CURRENT FILING DATE:  2003-09-05
;  PRIOR APPLICATION NUMBER: 09/079,030
;  PRIOR FILING DATE: 1998-05-14
;  PRIOR APPLICATION NUMBER: 08/874,807
;  PRIOR FILING DATE: 1997-06-13
;  NUMBER OF SEQ ID NOS: 229
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 229
;   LENGTH: 34
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-656-053B-229

  Query Match             50.0%;  Score 30;  DB 9;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||||||||||||||||||||||||||||
Db          4 CCTCGGGAAAGGGAAACCGAAACTGAAGCC 33


RESULT 14
US-10-363-345A-25757/c
; Sequence 25757, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 25757
;   LENGTH: 2308
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 25757
US-10-363-345A-25757

  Query Match             50.0%;  Score 30;  DB 8;  Length 2308;
  Best Local Similarity   72.2%;  Pred. No. 2.5;
  Matches   39;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||  ||||||||||| || ||       ||||  ||||||||||| || ||
Db       1222 GAAACCGAAACCGAAACGAAAACCGAAACCGAAACCGAAACCGAAACTAAAACC 1169
```

RESULT 15
US-10-363-345A-25758
; Sequence 25758, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 25758
; LENGTH: 2308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 25758
US-10-363-345A-25758

Query Match 50.0%; Score 30; DB 8; Length 2308;
Best Local Similarity 72.2%; Pred. No. 2.5;
Matches 39; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              ||||  ||||||||||  || ||      ||||  ||||||||||| || ||
Db       1087 GAAACCGAAACCGAAACGAAAACCGAAACCGAAACCGAAACCGAAACTAAAACC 1140
```

Search completed: March 30, 2006, 19:34:37
Job time : 454 secs

THIS PAGE BLANK (USPTO)

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).
Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2006, 19:27:13 ; Search time 376 Seconds
(without alignments)
637.181 Million cell updates/sec

Title: US-10-677-777-5
Perfect score: 60
Sequence: 1 ctcgggaaagggaaaccgaa.....ggaaaccgaaactgaagccc 60

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9263891 seqs, 1996499642 residues

Total number of hits satisfying chosen parameters: 18527782

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
8: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
11: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
12: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
13: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
15: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	27.4	45.7	575	6	US-09-925-065A-758549	Sequence 758549,
2	25.8	43.0	842	6	US-09-925-065A-77163	Sequence 77163, A
3	25.8	43.0	842	9	US-10-301-480-178402	Sequence 178402,
4	25.8	43.0	842	10	US-10-301-480-791811	Sequence 791811,
5	25.8	43.0	1315	11	US-11-096-568A-7223	Sequence 7223, Ap
6	25.6	42.7	884	14	US-11-000-463-25	Sequence 25, Appl
c 7	25.6	42.7	16001	14	US-11-004-762-5	Sequence 5, Appli
8	25.4	42.3	651	9	US-10-932-182A-5213	Sequence 5213, Ap
9	25.4	42.3	651	9	US-10-932-182A-5213	Sequence 5213, Ap
c 10	25.4	42.3	1534	6	US-09-925-065A-89353	Sequence 89353, A
c 11	25.4	42.3	1534	9	US-10-301-480-89513	Sequence 89513, A
c 12	25.4	42.3	1534	9	US-10-301-480-89514	Sequence 89514, A
c 13	25.4	42.3	1534	9	US-10-301-480-190594	Sequence 190594,
c 14	25.4	42.3	1534	10	US-10-301-480-702922	Sequence 702922,
c 15	25.4	42.3	1534	10	US-10-301-480-702923	Sequence 702923,
c 16	25.4	42.3	1534	10	US-10-301-480-804003	Sequence 804003,
c 17	25.4	42.3	1707	6	US-09-925-065A-694321	Sequence 694321,
18	25.4	42.3	1877	11	US-11-096-568A-12270	Sequence 12270, A
c 19	25.4	42.3	2272	6	US-09-925-065A-724382	Sequence 724382,
20	25.4	42.3	2286	14	US-11-136-527-231	Sequence 231, App
21	25.2	42.0	390	14	US-11-000-688-1185	Sequence 1185, Ap
22	25.2	42.0	591	10	US-10-301-480-448658	Sequence 448658,
23	25.2	42.0	591	10	US-10-301-480-448659	Sequence 448659,
24	25.2	42.0	591	10	US-10-301-480-1062067	Sequence 1062067,
25	25.2	42.0	591	10	US-10-301-480-1062068	Sequence 1062068,
26	25.2	42.0	593	6	US-09-925-065A-379418	Sequence 379418,
27	25.2	42.0	593	6	US-09-925-065A-379419	Sequence 379419,
c 28	25.2	42.0	983	10	US-10-301-480-567480	Sequence 567480,
c 29	25.2	42.0	983	10	US-10-301-480-1180889	Sequence 1180889,
30	25.2	42.0	1954	14	US-11-000-688-1187	Sequence 1187, Ap
31	25.2	42.0	2609	11	US-11-072-512-1365	Sequence 1365, Ap
c 32	24.6	41.0	333	9	US-10-932-182A-174483	Sequence 174483,
c 33	24.6	41.0	333	9	US-10-932-182A-174483	Sequence 174483,
34	24.6	41.0	3084	9	US-10-932-182A-2903	Sequence 2903, Ap
35	24.6	41.0	3084	9	US-10-932-182A-2903	Sequence 2903, Ap
36	24.4	40.7	86899	9	US-10-330-773-963	Sequence 963, App
37	24.2	40.3	591	6	US-09-925-065A-405024	Sequence 405024,
38	24.2	40.3	591	6	US-09-925-065A-405025	Sequence 405025,
39	24.2	40.3	593	10	US-10-301-480-471919	Sequence 471919,
40	24.2	40.3	593	10	US-10-301-480-471920	Sequence 471920,
41	24.2	40.3	593	10	US-10-301-480-1085328	Sequence 1085328,
42	24.2	40.3	593	10	US-10-301-480-1085329	Sequence 1085329,
c 43	24.2	40.3	995	10	US-10-301-480-594209	Sequence 594209,
c 44	24.2	40.3	995	10	US-10-301-480-1207618	Sequence 1207618,
c 45	24.2	40.3	2743	6	US-09-925-065A-7581	Sequence 7581, Ap

ALIGNMENTS

RESULT 1
US-09-925-065A-758549/c
; Sequence 758549, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 758549
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-758549

Query Match 45.7%; Score 27.4; DB 6; Length 575;
Best Local Similarity 69.8%; Pred. No. 5.5;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              |||||||||||  ||||| |||       ||||||||||||  |||  | |
Db        165 GGGAAAGGGAAAGGGAAACGGAAAGGGAAAGGGAAAGGGAAAAGGGAAGGGGA 113
```

RESULT 2
US-09-925-065A-77163
; Sequence 77163, Application US/09925065A

```
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 77163
;   LENGTH: 842
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-77163

  Query Match             43.0%;  Score 25.8;  DB 6;  Length 842;
  Best Local Similarity   67.9%;  Pred. No. 21;
  Matches   36;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              ||||||||||||  ||||  |||       | |||||  ||||  ||||  |||
Db        205 GGGAAAGGGAAAGGGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAA 257


RESULT 3
US-10-301-480-178402
; Sequence 178402, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 178402
;   LENGTH: 842
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-178402

  Query Match             43.0%;  Score 25.8;  DB 9;  Length 842;
  Best Local Similarity   67.9%;  Pred. No. 21;
  Matches   36;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              ||||||||||||  ||||  |||       | |||||  ||||  ||||  |||
Db        205 GGGAAAGGGAAAGGGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAA 257


RESULT 4
US-10-301-480-791811
; Sequence 791811, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 791811
;   LENGTH: 842
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-791811

  Query Match             43.0%;  Score 25.8;  DB 10;  Length 842;
  Best Local Similarity   67.9%;  Pred. No. 21;
  Matches   36;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          4 GGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAA 56
              ||||||||||||  ||||  |||       | |||||  ||||  ||||  |||
Db        205 GGGAAAGGGAAAGGGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAA 257


RESULT 5
US-11-096-568A-7223
; Sequence 7223, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
;  APPLICANT: Alexandrov, Nickolai et al.
;  TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
;  TITLE OF INVENTION:  Therby
;  FILE REFERENCE: 2750-1592PUS2
;  CURRENT APPLICATION NUMBER: US/11/096,568A
;  CURRENT FILING DATE:  2005-04-01
;  NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 7223
;   LENGTH: 1315
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1315)
;   OTHER INFORMATION: Ceres Seq. ID no. 15170907
US-11-096-568A-7223

  Query Match             43.0%;  Score 25.8;  DB 11;  Length 1315;
  Best Local Similarity   67.9%;  Pred. No. 23;
  Matches   36;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          6 GAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              |||   ||| | | | ||||||||    || ||||||||   ||||  |||||
Db        484 GAAGCAGAAGCTGCAGCTGAAGCCAAAAGGAAAAGGGAACTGGAAAGAGAAGC 536


RESULT 6
US-11-000-463-25
; Sequence 25, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Qian, Xiaohong B.
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Cao, Yi-Cheng
;  APPLICANT:  Drmanac, Radoje T.
```

```
;  TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
;  FILE REFERENCE: 785CIP4CN
;  CURRENT APPLICATION NUMBER: US/11/000,463
;  CURRENT FILING DATE:  2004-11-29
;  PRIOR APPLICATION NUMBER: 10/291,265
;  PRIOR FILING DATE: 2002-11-08
;  PRIOR APPLICATION NUMBER: PCT/US01/02623
;  PRIOR FILING DATE: 2001-01-25
;  PRIOR APPLICATION NUMBER: 09/922,279
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: 09/491,404
;  PRIOR FILING DATE: 2000-01-25
;  PRIOR APPLICATION NUMBER: 09/617,746
;  PRIOR FILING DATE: 2000-07-17
;  PRIOR APPLICATION NUMBER: 09/631,451
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 09/633,870
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 944
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 25
;   LENGTH: 884
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (137)..(715)
US-11-000-463-25

  Query Match             42.7%;  Score 25.6;  DB 14;  Length 884;
  Best Local Similarity   66.1%;  Pred. No. 25;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          3 CGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGC 58
              | ||| | |||||   | |||||    ||   ||||||||| | ||| ||||| |
Db        491 CAGGATATGGAAAAAAACCCTGAATTATCTAATGAAAGGGAACCTGAACCTGAACC 546


RESULT 7
US-11-004-762-5/c
; Sequence 5, Application US/11004762
; Publication No. US20060003953A1
; GENERAL INFORMATION:
;  APPLICANT: C. Frank Bennett
;  APPLICANT:  Madeline M. Butler
;  APPLICANT:  Nicholas M. Dean
;  APPLICANT:  Kenneth W. Dobie
;  APPLICANT:  Joshua Finger
;  APPLICANT:  Ravi Jain
;  APPLICANT:  Robert McKay
;  APPLICANT:  Brett P. Monia
;  APPLICANT:  Kathleen Myers
;  TITLE OF INVENTION: Compositions and their uses directed to bone growth modulators
;  FILE REFERENCE: BIOL0050US
;  CURRENT APPLICATION NUMBER: US/11/004,762
;  CURRENT FILING DATE:  2004-12-03
;  PRIOR APPLICATION NUMBER: US 60/527,370
;  PRIOR FILING DATE: 2003-12-04
;  PRIOR APPLICATION NUMBER: US 60/527,173
;  PRIOR FILING DATE: 2003-12-04
;  PRIOR APPLICATION NUMBER: US 60/527,172
;  PRIOR FILING DATE: 2003-12-04
;  PRIOR APPLICATION NUMBER: US 60/527,420
;  PRIOR FILING DATE: 2003-12-04
;  PRIOR APPLICATION NUMBER: US 60/527,174
;  PRIOR FILING DATE: 2003-12-04
;  PRIOR APPLICATION NUMBER: US 60/527,397
;  PRIOR FILING DATE: 2003-12-04
;  NUMBER OF SEQ ID NOS: 680
;  SOFTWARE: PatentSeq version 1.0
; SEQ ID NO 5
;   LENGTH: 16001
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-11-004-762-5

  Query Match             42.7%;  Score 25.6;  DB 14;  Length 16001;
  Best Local Similarity   66.1%;  Pred. No. 42;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          2 TCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | ||| ||||| ||     |    | |||||| || ||||| |||  ||||| || |
Db       7761 TGGGGGAAGGGCAAGGCCAGAAAAGGCCCCTGGGAAAAGGAAAAGGGAAACAGAGG 7706


RESULT 8
US-10-932-182A-5213
; Sequence 5213, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAO, YOSHIHIRO
;  APPLICANT:  NAKAMURA, NORIHISA
;  APPLICANT:  KODAMA, YUKIKO
;  APPLICANT:  FUJIMURA, TOMOKO
;  APPLICANT:  ASHIKARI, TOSHIHIKO
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
;  FILE REFERENCE: 030685-043
;  CURRENT APPLICATION NUMBER: US/10/932,182A
;  CURRENT FILING DATE:  2004-09-02
;  NUMBER OF SEQ ID NOS: 197023
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5213
;   LENGTH: 651
;   TYPE: DNA
;   ORGANISM: Saccharomyces pastorianus
US-10-932-182A-5213

  Query Match             42.3%;  Score 25.4;  DB 9;  Length 651;
  Best Local Similarity   64.4%;  Pred. No. 27;
  Matches   38;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              |||  ||||  ||||  |||||||   || |||  |||||| ||   || | |   |||
Db        371 CTCTAGAAAAAGAAAAGGAAACTGTTTCCTCTCAAGAAAGGAAAGAGGATATTATCGCC 429


RESULT 9
US-10-932-182A-5213
; Sequence 5213, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAO, YOSHIHIRO
;  APPLICANT:  NAKAMURA, NORIHISA
;  APPLICANT:  KODAMA, YUKIKO
;  APPLICANT:  FUJIMURA, TOMOKO
;  APPLICANT:  ASHIKARI, TOSHIHIKO
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
;  FILE REFERENCE: 030685-043
;  CURRENT APPLICATION NUMBER: US/10/932,182A
;  CURRENT FILING DATE:  2004-09-02
;  NUMBER OF SEQ ID NOS: 197023
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5213
;   LENGTH: 651
;   TYPE: DNA
;   ORGANISM: Saccharomyces pastorianus
US-10-932-182A-5213

  Query Match             42.3%;  Score 25.4;  DB 9;  Length 651;
  Best Local Similarity   64.4%;  Pred. No. 27;
  Matches   38;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          1 CTCGGGAAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAGCC 59
              |||  ||||  ||||  |||||||   || |||  |||||| ||   || | |   |||
Db        371 CTCTAGAAAAAGAAAAGGAAACTGTTTCCTCTCAAGAAAGGAAAGAGGATATTATCGCC 429
```

```
RESULT 10
US-09-925-065A-89353/c
; Sequence 89353, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 89353
;   LENGTH: 1534
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-89353

  Query Match             42.3%;  Score 25.4;  DB 6;  Length 1534;
  Best Local Similarity   68.6%;  Pred. No. 32;
  Matches   35;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          7 AAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | || ||||   |||  |||  | || ||||||||||| | ||||  | ||
Db        117 ACAGAGAAAAAAAAAGAGAACACTCTAGGGAAAGGGAACCAGAAAAAGTAG 67


RESULT 11
US-10-301-480-89513/c
; Sequence 89513, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 89513
;   LENGTH: 1534
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-89513

  Query Match             42.3%;  Score 25.4;  DB 9;  Length 1534;
  Best Local Similarity   68.6%;  Pred. No. 32;
  Matches   35;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          7 AAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | || ||||   |||  |||  | || ||||||||||| | ||||  | ||
Db        117 ACAGAGAAAAAAAAAGAGAACACTCTAGGGAAAGGGAACCAGAAAAAGTAG 67


RESULT 12
US-10-301-480-89514/c
; Sequence 89514, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 89514
;   LENGTH: 1534
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-89514

  Query Match             42.3%;  Score 25.4;  DB 9;  Length 1534;
  Best Local Similarity   68.6%;  Pred. No. 32;
  Matches   35;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          7 AAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | || ||||   |||  |||  | || ||||||||||| | ||||  | ||
Db        117 ACAGAGAAAAAAAAAGAGAACACTCTAGGGAAAGGGAACCAGAAAAAGTAG 67


RESULT 13
US-10-301-480-190594/c
; Sequence 190594, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190594
;   LENGTH: 1534
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-190594

  Query Match             42.3%;  Score 25.4;  DB 9;  Length 1534;
  Best Local Similarity   68.6%;  Pred. No. 32;
  Matches   35;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          7 AAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | || ||||   |||  |||  | || ||||||||||| | ||||  | ||
Db        117 ACAGAGAAAAAAAAAGAGAACACTCTAGGGAAAGGGAACCAGAAAAAGTAG 67


RESULT 14
US-10-301-480-702922/c
; Sequence 702922, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
```

; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 702922
; LENGTH: 1534
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-702922

Query Match 42.3%; Score 25.4; DB 10; Length 1534;
Best Local Similarity 68.6%; Pred. No. 32;
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy          7 AAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | || ||||   ||| ||| | || |||||||||||| | ||||  | ||
Db        117 ACAGAGAAAAAAAAAAGAGAACACTCTAGGGAAAGGGAACCAGAAAAAGTAG 67
```

RESULT 15
US-10-301-480-702923/c
; Sequence 702923, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 702923
; LENGTH: 1534
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-702923

Query Match 42.3%; Score 25.4; DB 10; Length 1534;
Best Local Similarity 68.6%; Pred. No. 32;
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy          7 AAAGGGAAACCGAAACTGAAGCCCCTCGGGAAAGGGAAACCGAAACTGAAG 57
              | || ||||   ||| ||| | || |||||||||||| | ||||  | ||
Db        117 ACAGAGAAAAAAAAAAGAGAACACTCTAGGGAAAGGGAACCAGAAAAAGTAG 67
```

Search completed: March 30, 2006, 19:40:59
Job time : 377 secs

THIS PAGE BLANK (USPTO)